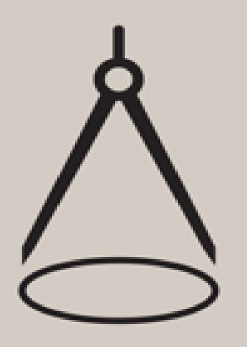
Autor bestseller de *The New York Times*

STEVEN PINKER



Racionalidad

Qué es, por qué escasea y cómo promoverla

PAIDÓS

Índice

SINOPSIS

En el siglo XXI la humanidad está alcanzando nuevas cotas en la comprensión científica y, al mismo tiempo, parece estar enloqueciendo. ¿Cómo puede una especie que ha desarrollado vacunas para la COVID-19 en menos de un año producir tantas noticias falsas, remedios de curanderos y teorías de la conspiración?

Pinker rechaza el cínico lugar común de que somos una especie irracional: cavernícolas extemporáneos cargados de sesgos, falacias e ilusiones. Después de todo, hemos descubierto las leyes de la naturaleza, hemos prolongado y enriquecido nuestras vidas, y hemos fijado los puntos de referencia de la racionalidad misma. En lugar de ello, explica que pensamos de formas sensatas en los contextos de baja tecnología en los que transcurre la mayor parte de nuestras vidas, pero no sacamos partido de las potentes herramientas del razonamiento que hemos construido a lo largo de los milenios: la lógica, el pensamiento crítico, la probabilidad, la correlación y la causalidad, y la toma de decisiones. Estas herramientas no son un componente estándar de nuestros currículos educativos y, hasta la fecha, nunca se habían presentado de un modo claro y ameno.

Racionalidad explora asimismo su contraria: cómo la búsqueda racional del interés propio, la solidaridad sectaria y la mitología edificante de los individuos puede desembocar en la irracionalidad paralizante en una sociedad. La racionalidad colectiva depende de normas explícitamente diseñadas para promover la objetividad y la verdad.

La racionalidad es importante. Nos permite tomar mejores decisiones en nuestra vida y en la esfera pública, y es el principal motor de la justicia social y del progreso moral. Rebosante de perspicacia y de humor, Racionalidad nos iluminará, nos inspirará y nos empoderará.

STEVEN PINKER

RACIONALIDAD

Qué es
Por qué parece escasear
Por qué es importante

Traducción de Pablo Hermida Lazcano

PAIDÓS Contextos



A Roslyn Wiesenfeld Pinker

¿Qué fuera el hombre si su bien más alto,
el logro de su vida consistiese
solo en dormir y sustentar su cuerpo?
Un bruto nada más. El que nos hizo,
dotados de talento tan preclaro,
que vemos lo pasado y lo futuro,
por cierto no nos dio razón divina,
tal aptitud a fin de que en nosotros
enmoheciese por falta de uso.

HAMLET

Prefacio

La racionalidad debería ser la estrella polar de todo cuanto pensamos y hacemos (si no estás de acuerdo con esta afirmación, ¿son racionales tus objeciones?). Sin embargo, en una época pródiga en recursos inéditos para el razonamiento, la esfera pública está infestada de *fake news* o noticias falsas, remedios de charlatanes, teorías de la conspiración y retórica de la «posverdad».

¿Cómo podemos llegar a discernir el sentido del sinsentido? La pregunta es urgente. En la tercera década del tercer milenio nos enfrentamos a amenazas mortales para nuestra salud, nuestra democracia y la habitabilidad de nuestro planeta. Si bien los problemas son abrumadores, existen soluciones y nuestra especie posee los recursos intelectuales para hallarlas. No obstante, entre nuestros más feroces problemas actuales figura la dificultad de convencer a la gente para que acepte las soluciones cuando efectivamente las encontramos.

Millares de comentarios han lamentado nuestro déficit de razón y se ha popularizado la creencia de que las personas somos sencillamente irracionales. En las ciencias sociales y en los medios de comunicación se retrata al ser humano como un cavernícola extemporáneo, preparado para reaccionar ante un león en la hierba con una serie de sesgos, puntos ciegos, falacias e ilusiones (la entrada de Wikipedia para los sesgos cognitivos enumera casi dos centenares).

Ahora bien, como científico cognitivo, yo no puedo aceptar la cínica visión de que el cerebro humano es un capazo de ilusiones. Los cazadores recolectores —nuestros ancestros y contemporáneos— no son conejos nerviosos, sino cerebrales solucionadores de problemas. Un listado de las modalidades de nuestra estupidez no basta para explicar por qué somos tan inteligentes: cómo hemos descubierto las leyes de la naturaleza,

transformado el planeta, prolongado y enriquecido nuestras vidas y, en particular, articulado las reglas de la racionalidad que incumplimos con tanta frecuencia.

Desde luego, soy de los primeros en insistir en que solo podemos comprender la naturaleza humana considerando el desajuste entre el entorno en el que hemos evolucionado y el entorno en el que nos hallamos en la actualidad. Pero el mundo al que nuestra mente está adaptada no es únicamente la sabana del Pleistoceno. Es cualquier medio no tecnocrático y no académico —lo cual equivale a la mayor parte de la experiencia humana— en el que los modernos instrumentos de la racionalidad, como las fórmulas estadísticas y los conjuntos de datos, no se encuentran disponibles o son inaplicables. Como veremos, cuando a las personas se les plantean problemas más próximos a su realidad vivida y enmarcados en sus formas naturales de encontrarse con el mundo, no son tan estúpidas como parecen. No es que esto nos saque del apuro, pero lo cierto es que hoy en día disponemos de refinados instrumentos de la razón y que, tanto individual como colectivamente, salimos ganando cuando los entendemos y los aplicamos.

Este libro nació de un curso impartido en Harvard, que exploraba la naturaleza de la racionalidad y el enigma de por qué esta parece escasear tanto. Al igual que a muchos psicólogos, a mí me encanta enseñar los deslumbrantes descubrimientos, galardonados con el Premio Nobel, sobre las enfermedades que afligen a la razón humana, y considero que estos figuran entre los regalos más profundos para el conocimiento de los que nuestra ciencia puede jactarse. Y, como tantos otros, creo que las cotas de la racionalidad que con tanta frecuencia no somos capaces de alcanzar deberían constituir un objetivo de la educación y de la ciencia popular. Del mismo modo que los ciudadanos deberían comprender los principios básicos de la historia, la ciencia y la palabra escrita, tendrían que dominar las herramientas intelectuales del razonamiento certero. Estas incluyen la lógica, el pensamiento crítico, la probabilidad, la correlación y la causalidad, las maneras óptimas de ajustar nuestras creencias y comprometernos con decisiones con pruebas inciertas, y los criterios para tomar decisiones racionales tanto en solitario como con otras personas.

Estas herramientas del razonamiento resultan indispensables a la hora de evitar la estupidez tanto en nuestra vida personal como en las políticas públicas. Nos ayudan a calibrar las decisiones arriesgadas, a evaluar las afirmaciones dudosas, a entender las paradojas desconcertantes y a comprender mejor las vicisitudes y las tragedias de la vida. No obstante, no he conocido ningún libro que tratase de explicar todos estos instrumentos.

La otra inspiración para este libro fue mi constatación de que, pese a toda su fascinación, el currículo de Psicología Cognitiva me había dejado mal equipado para responder las preguntas que me hacían con más frecuencia cuando contaba que estaba impartiendo un curso sobre la racionalidad. ¿Por qué cree la gente que Hillary Clinton dirigía una red de pedofilia desde una pizzería, o que las estelas de vapor de los aviones son drogas que alteran la mente, rociadas por un programa gubernamental secreto? Los temas estelares de mis clases típicas, como la falacia del jugador y la falacia de la tasa base, no arrojaban mucha luz sobre los enigmas que están convirtiendo la irracionalidad humana en una cuestión tan apremiante en nuestros días. Esos enigmas me arrastraron hasta nuevos territorios, como la naturaleza del rumor, la sabiduría popular y el pensamiento conspirativo; el contraste entre la racionalidad del individuo y de la comunidad; y la distinción entre dos modalidades de creencia: la mentalidad realista y la mentalidad mitológica.

Finalmente, aunque pueda antojarse paradójico exponer argumentos racionales a favor de la propia racionalidad, esta es una tarea oportuna. Algunos incurren en la paradoja opuesta, citando razones (presumiblemente racionales, ¿o por qué habríamos de prestarles atención si no?) por las que la racionalidad está sobrevalorada: que las personalidades lógicas son tristes y están reprimidas, que el pensamiento analítico debe subordinarse a la justicia social, y un buen corazón y una intuición fiable son rutas más seguras hacia el bienestar que la lógica tenaz y los razonamientos. Son muchos los que actúan como si la racionalidad estuviese obsoleta: como si el objetivo de la argumentación fuese desacreditar a nuestros adversarios más que razonar colectivamente en pro de las creencias más defendibles. En

una época en la que la racionalidad parece al mismo tiempo más amenazada y más esencial que nunca, *Racionalidad* es, por encima de todo, una afirmación de la racionalidad.

UN TEMA FUNDAMENTAL de este libro es que ninguno de nosotros es lo suficientemente racional como para llegar sistemáticamente a conclusiones sólidas: la racionalidad emerge de una comunidad de razonadores que detectan mutuamente las falacias ajenas. Con ese espíritu doy las gracias a los razonadores que han tornado más racional este libro. Ken Binmore, Rebecca Newberger Goldstein, Gary King, Jason Nemirow, Roslyn Pinker, Keith Stanovich y Martina Wiese comentaron de manera incisiva el primer borrador. Charleen Adams, Robert Aumann, Joshua Hartshorne, Louis Liebenberg, Colin McGinn, Barbara Mellers, Hugo Mercier, Judea Pearl, David Ropeik, Michael Shermer, Susanna Siegel, Barbara Spellman, Lawrence Summers, Philip Tetlock y Juliani Vidal revisaron capítulos de sus respectivas áreas de especialización. Muchas fueron las preguntas que surgieron mientras planeaba y escribía el libro, y estas fueron respondidas por Daniel Dennett, Emily-Rose Eastop, Baruch Fischhoff, Reid Hastie, Nathan Kuncel, Ellen Langer, Jennifer Lerner, Beau Lotto, Daniel Loxton, Gary Marcus, Philip Maymin, Don Moore, David Myers, Robert Proctor, Fred Shapiro, Mattie Toma, Jeffrey Watumull, Jeremy Wolfe y Steven Zipperstein. Para la transcripción, la verificación de hechos y la búsqueda de referencias he contado con la pericia de Mila Bertolo, Martina Wiese y Kai Sandbrink, y asimismo me he beneficiado del análisis de datos originales de Bertolo, Toma y Julian De Freitas. Agradezco también las preguntas y las sugerencias de los estudiantes y profesores del curso de Educación General 1066: Racionalidad, especialmente de Mattie Toma y Jason Nemirow.

Mi especial gratitud para mi sabia y alentadora editora, Wendy Wolf, por trabajar conmigo en este libro, nuestro sexto, y para Katya Rice, por corregir nuestro noveno; también para mi agente literario, John Brockman, por su estímulo y sus consejos en nuestro noveno. Agradezco asimismo el

apoyo a lo largo de muchos años de Thomas Penn, Pen Vogler y Stefan McGrath, de Penguin UK. Ilavenil Subbiah ha diseñado una vez más los gráficos, y le agradezco su labor y su aliento.

Rebecca Newberger Goldstein ha desempeñado un papel especial en la concepción de estas páginas, pues ha sido ella quien me ha inculcado que el realismo y la razón son ideales que han de ser destacados y defendidos. Deseo expresar asimismo mi amor y mi agradecimiento a los demás miembros de mi familia: Yael y Solly; Danielle; Rob, Jack y David; Susan, Martin, Eva, Carl y Eric; y a mi madre Roslyn, a quien está dedicado este libro.

RACIONALIDAD

CAPÍTULO 1

¿Cuán racional es este animal?

El hombre es un animal racional. Eso es al menos lo que nos han contado. En el transcurso de mi larga vida he buscado diligentemente pruebas en favor de esta afirmación, pero hasta ahora no he tenido la fortuna de toparme con ellas.

BERTRAND RUSSELL[1]

Aquel que es capaz de criticar con mayor elocuencia y agudeza la debilidad de la mente humana es considerado casi divino por sus compañeros.

BARUCH SPINOZA[2]

Homo sapiens significa «homínido sabio» y, en muchos sentidos, nos hemos ganado el epíteto específico de nuestro binomio linneano. Nuestra especie ha datado el origen del universo, ha sondeado la naturaleza de la materia y la energía, ha descifrado los secretos de la vida, ha desentrañado los circuitos de la conciencia y ha hecho una crónica de nuestra historia y nuestra diversidad. Hemos aplicado estos conocimientos a potenciar nuestro florecimiento, mitigando los flagelos que empobrecían a nuestros antepasados durante la mayor parte de nuestra existencia. Hemos aplazado nuestra esperada cita con la muerte desde los treinta años hasta más de setenta (ochenta en los países desarrollados), hemos reducido la pobreza extrema del 95 % de la humanidad a menos del 9 %, hemos disminuido veinte veces las tasas de mortalidad por las guerras y cien veces las muertes provocadas por la hambruna.[3] Incluso cuando la antigua maldición de la peste ha resurgido en el siglo XXI, hemos identificado las causas en cuestión de días, hemos secuenciado su genoma en unas semanas y hemos administrado vacunas en un año, manteniendo su número de víctimas en una fracción de las de otras pandemias históricas.

Los recursos cognitivos para entender el mundo y someterlo a nuestra voluntad no son un trofeo de la civilización occidental; son el patrimonio de nuestra especie. Los sans del desierto del Kalahari en el sur de África son uno de los pueblos más antiguos del mundo y su estilo de vida, basado en la búsqueda y recolección de alimentos, mantenido hasta fechas recientes, nos permite vislumbrar las formas en las que ha transcurrido la mayor parte de nuestra existencia como humanos.[4] Los cazadores recolectores no se limitan a arrojar lanzas a los animales que pasan o a coger las frutas y los frutos secos que crecen a su alrededor.[5] El experto en rastreo Louis Liebenberg, que ha trabajado durante décadas con los sans, ha descrito cómo estos deben su supervivencia a una mentalidad científica.[6] Razonan a partir de datos fragmentarios hasta llegar a conclusiones remotas con un manejo intuitivo de la lógica, el pensamiento crítico, el razonamiento estadístico, la inferencia causal y la teoría de juegos.

Los sans practican la caza de persistencia, que explota nuestras tres características más conspicuas: nuestro bipedismo, que nos permite correr eficientemente; nuestra falta de pelo corporal, que nos permite liberar el calor en los climas cálidos, y nuestra gran cabeza, que nos permite ser racionales. Los sans despliegan esta racionalidad para rastrear a los animales que huyen a partir de las huellas de sus pezuñas, sus efluvios y otras pistas, persiguiéndolos hasta que estos se desploman por agotamiento e insolación.[7] A veces, los sans rastrean a un animal a lo largo de uno de sus caminos habituales o, cuando el rastro se pierde, buscando en círculos crecientes en torno a las últimas huellas conocidas. Pero con frecuencia los rastrean mediante el razonamiento.

Los cazadores distinguen docenas de especies por las formas y la separación de sus huellas, ayudados por su comprensión de las causas y los efectos. Pueden inferir que una huella profundamente puntiaguda procede de una ágil gacela saltarina, que necesita un buen agarre, mientras que una huella de pies planos procede de un pesado kudú, que tiene que soportar su peso. Son capaces de determinar el sexo de los animales a partir de la configuración de sus huellas y de la ubicación relativa de su orina respecto de sus extremidades posteriores y sus excrementos. Usan estas categorías para hacer deducciones silogísticas: el racífero común y el duiker pueden cazarse en la estación lluviosa porque la arena mojada los obliga a abrir sus pezuñas y entumece sus articulaciones; el kudú y el eland pueden atraparse en la temporada seca porque se cansan fácilmente en la arena suelta. Es la estación seca y el animal que ha dejado estas huellas es un kudú; por consiguiente, este animal puede ser cazado.

Los sans no solo clasifican los animales en categorías, sino que hacen asimismo distinciones lógicas más sutiles. Reconocen a los individuos dentro de una especie por las huellas de sus pezuñas, variaciones y rasguños reveladores. Y distinguen las características permanentes de un individuo, como su especie y su sexo, de las condiciones transitorias, como la fatiga, que infieren de los signos del arrastre de las pezuñas y las paradas para descansar. Desafiando la patraña de que los pueblos premodernos carecen de concepto de *tiempo*, estiman la edad de un animal a partir del tamaño y la nitidez de sus huellas, y pueden datar su rastro en función de lo recientes

que sean las huellas, la humedad de la saliva o los excrementos, el ángulo del sol respecto de un lugar de descanso sombreado y el palimpsesto de huellas superpuestas de otros animales. La caza por persistencia no podría tener éxito sin esas sutilezas lógicas. Un cazador no puede rastrear cualquier órice de entre los muchos que han dejado huellas, sino solo el que ha estado persiguiendo hasta la extenuación.

Los sans cultivan asimismo el pensamiento crítico. Saben no fiarse de sus primeras impresiones y aprecian los peligros de ver aquello que desean ver. Tampoco aceptan los argumentos de autoridad: cualquiera, incluido un joven osado, puede echar por tierra una conjetura o proponer la suya propia hasta que de la disputa surja un consenso. Aunque son principalmente los hombres los que se dedican a la caza, las mujeres son tan expertas como ellos en la interpretación de los rastros, y Liebenberg refiere que una joven llamada !Nasi «puso en evidencia a los hombres».[8]

Los sans adaptan su creencia en una hipótesis en función de lo diagnósticas que sean las pruebas, una cuestión de probabilidad condicional. Un pie de puercoespín, por ejemplo, tiene dos almohadillas proximales, en tanto que el tejón de la miel tiene una solo, pero puede que una almohadilla no deje huella en un suelo duro. Esto significa que, aunque es alta la probabilidad de que un rastro tenga una huella de almohadilla, dado que fue dejado por un tejón de la miel, la probabilidad inversa, que un rastro fuese dejado por un tejón de la miel dado que tiene una huella de almohadilla, es más baja (ya que también podría tratarse de una huella incompleta de un puercoespín). Los sans no confunden estas probabilidades condicionales: saben que, como dos huellas de almohadillas solo podrían haber sido dejadas por un puercoespín, la probabilidad de un puercoespín, dadas dos huellas de almohadillas, es alta.

Los sans calibran también su creencia en una hipótesis conforme a la plausibilidad previa de esta. Si las huellas son ambiguas, asumirán que proceden de una especie común; solamente si las pruebas son definitivas concluirán que provienen de una más rara.[9] Como veremos, esa es la esencia del razonamiento bayesiano.

Otra facultad crítica ejercida por los sans es la distinción entre causalidad y correlación. Liebenberg recuerda: «Un rastreador, Boroh//xao, me contó que, cuando el pájaro canta, seca la tierra y hace que las raíces sean buenas para comer. Después, !Nate y /Uase me dijeron que Boroh//xao estaba equivocado: no es el *pájaro* el que seca la tierra, es el *sol* el que la seca. El pájaro solo les está *diciendo* que la tierra se secará en los meses siguientes y que es la época del año en la que las raíces son buenas para comer».[10]

Los sans utilizan su conocimiento de la textura causal de su entorno no solo para entender cómo es este, sino también para imaginar cómo podría ser. Al representar escenarios en su imaginación, pueden pensar varios pasos por delante de los animales en su mundo e idear intrincadas trampas para atraparlos. Se ancla al suelo un extremo de una rama elástica y se dobla el palo por la mitad; en el otro, se ata un lazo camuflado con ramillas y arena, sujeto mediante un gatillo. Colocan las trampas en las aberturas de las barreras que han construido en torno al lugar de descanso de un antílope y guían al animal hacia el sitio letal con un obstáculo que el antílope debe superar. O bien atraen a un avestruz hacia una trampa localizando sus huellas bajo una acacia erioloba o espina de camello (cuyas vainas son un manjar para los avestruces), dejando bien visible un hueso demasiado grande para ser tragado por el avestruz, que atrae su atención hacia otro hueso más pequeño, pero todavía intragable, que conduce hasta un hueso más pequeño, el cebo en la trampa.

No obstante, pese a la eficacia mortífera de la tecnología de los sans, estos han sobrevivido en un desierto implacable durante más de cien mil años sin exterminar a los animales de los que dependen. Durante una sequía, piensan por anticipado en lo que sucedería si mataran la última planta o animal de su especie, y perdonan la vida a los miembros de las especies amenazadas.[11] Adaptan sus planes de conservación a las diferentes vulnerabilidades de las plantas, que no pueden migrar, pero que se recuperan con rapidez cuando vuelven las lluvias, y de los animales, que pueden sobrevivir a una sequía pero tardan mucho tiempo en restablecer su número de individuos. E imponen estos esfuerzos de conservación contra la tentación constante de la caza furtiva (pues todos creen que deberían

explotar las especies raras porque, si no lo hacen ellos, lo harán los demás), con una extensión de las normas de reciprocidad y bienestar colectivo que gobiernan todos sus recursos. Para un cazador san resulta impensable no compartir la carne con un compañero del grupo que tiene las manos vacías, o excluir a un grupo vecino expulsado de su territorio azotado por la sequía, pues saben que los recuerdos perduran y algún día pueden volverse las tornas.

La sapiencia de los sans acentúa la paradoja de la racionalidad humana. Pese a nuestra antigua capacidad para razonar, hoy estamos inundados de recordatorios de las falacias y los disparates de nuestros semejantes. La gente apuesta y juega a la lotería, donde tiene garantizadas las pérdidas, y no consigue invertir en su jubilación, donde tiene garantizadas las ganancias. Tres cuartas partes de los estadounidenses creen al menos en un fenómeno que desafía las leyes de la ciencia, incluidos la sanación psíquica (55 %), la percepción extrasensorial (41 %), las casas encantadas (37 %) y los fantasmas (32 %), lo cual significa, además, que algunas personas creen en las casas encantadas por los fantasmas sin creer en los fantasmas.[12] En los medios sociales, las *fake news* (tales como «Joe Biden llama a los partidarios de Trump "la escoria de la sociedad"» y «Hombre arrestado en Florida por sedar y violar caimanes en los Everglades») se difunden más lejos y más rápido que la verdad, y los humanos tienen más probabilidades de propagarlas que los bots.[13]

Ha llegado a ser un lugar común concluir que los humanos son sencillamente irracionales, más parecidos a Homer Simpson que al señor Spock; más a Alfred E. Neuman[*] que a John von Neumann. Y, continúan los cínicos, ¿qué otra cosa cabría esperar de los descendientes de los cazadores recolectores cuya mente fue seleccionada para evitar convertirse en almuerzo de los leopardos? Pero los psicólogos evolucionistas, conscientes del ingenio de los pueblos cazadores recolectores, insisten en que los humanos evolucionaron para ocupar el «nicho cognitivo»: la capacidad de aventajar a la naturaleza con el lenguaje, la sociabilidad y el conocimiento.[14] Si los humanos contemporáneos parecen irracionales, no culpemos a los cazadores recolectores.

¿Cómo podemos entender entonces esta cosa llamada *racionalidad*, que parecería ser nuestro derecho de nacimiento, pero que es flagrantemente desdeñada con tanta frecuencia? El punto de partida consiste en apreciar que la racionalidad no es un poder que un agente posee o no posee, como la visión de rayos X de Superman. Es un juego de herramientas que puede alcanzar objetivos particulares en mundos particulares. Para comprender qué es la racionalidad, por qué parece escasear y por qué es importante, hemos de comenzar con las verdades fundamentales de la racionalidad misma: las formas en las que *debería* razonar un agente inteligente, dados sus objetivos y el mundo en el que vive. Estos modelos «normativos» provienen de la lógica, la filosofía, las matemáticas y la inteligencia artificial, y suponen la mejor comprensión de la solución «correcta» de un problema y del modo de hallarla por nuestra parte. Sirven de aspiración para aquellos que desean ser racionales, que deberíamos ser todos. Un objetivo primordial de este libro es explicar las herramientas normativas de la razón más ampliamente aplicables; estas constituyen los temas de los capítulos 3 a 9.

Los modelos normativos sirven asimismo como puntos de referencia para evaluar cómo razonamos de hecho los torpes e incompetentes humanos, el tema de la psicología y de las demás ciencias del comportamiento. Las múltiples formas en las que las personas ordinarias no alcanzan estos puntos de referencia se han vuelto famosas gracias a las investigaciones, galardonadas con el Premio Nobel, de Daniel Kahneman, Amos Tversky y otros psicólogos y economistas comportamentales.[15] Cuando los juicios de las personas se desvían de un modelo normativo, como sucede con tanta frecuencia, tenemos un enigma que resolver. A veces, la disparidad revela una irracionalidad genuina: el cerebro humano no es capaz de hacer frente a la complejidad de un problema, o arrastra un error que lo conduce obstinadamente una y otra vez hacia la respuesta incorrecta.

Pero en muchos casos la aparente locura de los individuos tiene su explicación. Puede que se les haya presentado un problema en un formato engañoso y, cuando este se traduce a un lenguaje más amigable para la mente, lo resuelven. O puede que el propio modelo normativo sea correcto

únicamente en un entorno particular, y que las personas perciban acertadamente que no están en él, por lo que el modelo no resulta aplicable. O bien puede que el modelo esté diseñado para lograr un objetivo determinado y, para bien o para mal, los individuos anden tras una meta diferente. En los próximos capítulos veremos ejemplos de todas estas circunstancias atenuantes. En el penúltimo capítulo se expondrá cómo algunos de los floridos estallidos de irracionalidad actuales pueden interpretarse como la persecución racional de objetivos distintos de una comprensión objetiva del mundo.

Aunque las explicaciones de la irracionalidad pueden absolver a las personas del cargo de absoluta estupidez, comprender no equivale a perdonar. A veces, podemos tener unas expectativas más elevadas para las personas. Podemos enseñarlas a identificar un problema profundo bajo sus disfraces superficiales. Podemos incitarlas a aplicar sus mejores hábitos de pensamiento fuera de sus zonas de confort. Y podemos inspirarlas para poner sus miras por encima de los objetivos contraproducentes o colectivamente destructivos. Estas son otras de las aspiraciones de este libro.

Dado que una idea recurrente en el estudio del juicio y la toma de decisiones es que los humanos devienen más racionales cuando la información que manejan es más vívida y relevante, permítaseme pasar a los ejemplos. Cada uno de estos clásicos —de las matemáticas, la lógica, la probabilidad y la predicción— revela una peculiaridad de nuestro razonamiento y servirá de anticipo de los estándares normativos de la racionalidad (y de las formas en que la gente se aparta de ellos) de los capítulos siguientes.

TRES SENCILLOS PROBLEMAS DE MATEMÁTICAS

Todo el mundo recuerda los tormentos del instituto con los problemas de álgebra, en los que se pedía calcular dónde se encontraría el tren que salía de Eastford hacia el oeste a ciento diez kilómetros por hora con el tren

que salía de Westford, a cuatrocientos veinte kilómetros de distancia, y que viajaba hacia el este a noventa y seis kilómetros por hora. Estos tres son más sencillos y puedes resolverlos de cabeza:

- Un *smartphone* y una funda cuestan ciento diez dólares en total. El teléfono cuesta cien dólares más que la funda. ¿Cuánto cuesta la funda?
- Se necesitan ocho impresoras durante ocho minutos para imprimir ocho folletos. ¿Cuánto tardarían veinticuatro impresoras en imprimir veinticuatro folletos?
- En un campo hay una zona de maleza. Cada día, esa zona duplica su tamaño. El área tarda treinta días en cubrir el campo entero. ¿Cuánto tiempo tardó en cubrir la mitad del campo?

La respuesta al primer problema es cinco dólares. Si eres como la mayoría de la gente, habrás respondido diez dólares. Pero si esa fuera la respuesta correcta, el teléfono costaría ciento diez dólares (cien dólares más que la funda) y el precio total de ambos artículos sería ciento veinte dólares.

La respuesta a la segunda pregunta es ocho minutos. Una impresora tarda ocho minutos en imprimir un folleto, luego, como hay tantas impresoras como folletos y están trabajando simultáneamente, el tiempo de imprimir los folletos es el mismo.

La respuesta al tercer problema es veintinueve días. Si la zona de maleza duplica su tamaño cada día, entonces, retrocediendo desde el día en que el campo quedó completamente cubierto, debió de estar medio cubierto el día anterior.

El economista Shane Frederick pasó estas preguntas (con diferentes ejemplos) a miles de estudiantes universitarios y descubrió que cinco de cada seis respondían mal al menos una de ellas, en tanto que uno de cada tres respondía *todas* mal.[16] Sin embargo, cada pregunta tiene una respuesta sencilla que casi todo el mundo entiende cuando se le explica. La dificultad estriba en que la gente desvía la atención hacia aspectos superficiales del problema, que consideran erróneamente relevantes para la respuesta, como

los números redondos cien y diez en el primer problema y el hecho de que el número de impresoras sea el mismo que el número de minutos en el segundo.

Frederick da a su batería de baja tecnología el nombre de *test de reflexión cognitiva* y sugiere que este pone de manifiesto una escisión entre dos sistemas cognitivos, que más tarde haría famosos Kahneman (coautor de algunos de sus trabajos) en el superventas de 2011 *Thinking Fast and Slow (Pensar rápido, pensar despacio)*. El sistema 1 opera rápidamente y sin ningún esfuerzo, y nos seduce con las respuestas incorrectas; el sistema 2 requiere concentración, motivación y la aplicación de reglas aprendidas, y nos permite comprender las correctas. Nadie piensa que se trate literalmente de dos sistemas anatómicos del cerebro; son dos modos de operación que involucran múltiples estructuras cerebrales. El sistema 1 implica juicios instantáneos; el sistema 2 implica pensárselo dos veces.

La lección del test de reflexión cognitiva es que los errores garrafales de razonamiento pueden ser fruto de la irreflexión más que de la ineptitud.[17] Incluso los estudiantes del Instituto de Tecnología de Massachusetts, sobresaliente en matemáticas, resolvieron correctamente por término medio dos de los tres problemas. Como cabía esperar, los resultados están correlacionados con la destreza en matemáticas, pero también con la paciencia. Los individuos que se describen a sí mismos como no impulsivos y que prefieren esperar un pago más grande en un mes que recibir uno más pequeño inmediatamente, tienen menos probabilidades de caer en trampas.[18]

Los dos primeros problemas parecen preguntas capciosas. Ello se debe al hecho de que proporcionan detalles que, en el ir y venir de la conversación, resultarían relevantes para lo que el hablante está preguntando, pero que en estos ejemplos está diseñado para llevar por mal camino al oyente (los resultados mejoran cuando el *smartphone* cuesta, pongamos por caso, setenta y tres dólares más que la funda y la suma asciende a ochenta y nueve dólares).[19] Pero ni que decir tiene que en la vida real también existen cebos en forma de embaucamientos y cantos de sirena que nos apartan de las buenas decisiones, y ser racional consiste en parte en resistirse a ellos. Quienes se dejan engañar por las respuestas seductoras

pero falsas en el test de reflexión cognitiva parecen ser menos racionales en otros ámbitos, como a la hora de rechazar ofertas lucrativas que requieren un poco de espera o algún riesgo.

Y el tercer problema, el de la zona de maleza, no es una pregunta con trampa, sino que explota una auténtica debilidad cognitiva. La intuición humana no capta el crecimiento exponencial (geométrico), a saber, algo que aumenta a un ritmo creciente, proporcional a su tamaño actual, como el interés compuesto, el crecimiento económico y la propagación de una enfermedad contagiosa.[20] La gente lo confunde con el aumento constante o la ligera aceleración, y su imaginación no sigue el ritmo de la duplicación incesante. Si depositas 400 dólares al mes en una cuenta de pensiones que gana un 10 % anual, ¿a cuánto ascenderán tus ahorros al cabo de 40 años? Muchas personas calculan en torno a 200.000 dólares, que es lo que obtenemos multiplicando 400 por 12 por el 110 % por 40. Algunos saben que esa cifra no puede ser correcta y ajusta al alza sus cálculos, pero nunca lo suficiente. Casi nadie llega a la respuesta correcta: 2,5 *millones* de dólares. Se ha descubierto que las personas con una comprensión más precaria del crecimiento exponencial ahorran menos para su jubilación y se endeudan más con las tarjetas de crédito, dos caminos hacia la miseria.[21]

La incapacidad de visualizar el despegue exponencial puede hacer caer en la trampa también a los expertos, incluidos los expertos en sesgos cognitivos. Cuando la COVID-19 llegó a Estados Unidos y a Europa en febrero de 2020, varios científicos sociales (incluidos dos héroes de este libro, aunque no el propio Kahneman) opinaron que la gente estaba sucumbiendo a un pánico irracional, porque había leído acerca de un par de casos espantosos y se dejaba llevar por el «sesgo de la disponibilidad» y el «descuido de la probabilidad». El riesgo objetivo en ese momento, observaban, era menor que el de la gripe o el de la faringitis estreptocócica, que todo el mundo acepta con tranquilidad.[22] La falacia de la reprimenda por la falacia consistía en subestimar el ritmo acelerado al que podía propagarse una enfermedad tan contagiosa como la COVID, en la que cada paciente no solo infecta a otros nuevos, sino que convierte a cada uno de ellos en un contagiador. La única muerte confirmada en Estados Unidos el 1 de marzo creció en las semanas sucesivas a 2, 5, 40, 264, 901 y 1.729

muertes diarias, ascendiendo a más de 100.000 a la altura del 1 de junio y convirtiéndose rápidamente en la principal causa de mortalidad del país.[23] Por supuesto, los autores de esos oscuros artículos de opinión no pueden ser culpados de la despreocupación que llevó a tantos dirigentes y ciudadanos a caer en una peligrosa complacencia, pero sus comentarios muestran cuán profundamente arraigados pueden estar los sesgos cognitivos.

¿Por qué *malsubestima* la gente el crecimiento exponencial, como podría haber dicho George W. Bush?* Siguiendo la tradición del médico de la obra teatral de Molière, que explicaba que el opio provoca el sueño en virtud de su poder dormitivo, los científicos sociales atribuyen los errores a un «sesgo del crecimiento exponencial». Con menos circularidad, podríamos apuntar a la fugacidad de los procesos exponenciales en los entornos naturales (con anterioridad a innovaciones históricas tales como el crecimiento económico y el interés compuesto). Las cosas que no pueden durar para siempre no lo hacen, y los organismos solo pueden multiplicarse hasta el momento en que agotan, contaminan o saturan sus entornos, doblegando la curva exponencial para convertirla en una S. Esto incluye las pandemias, que desaparecen una vez que un número suficiente de huéspedes susceptibles del rebaño han muerto o han desarrollado la inmunidad.

UN SENCILLO PROBLEMA LÓGICO

Si algo mora en el corazón de la racionalidad, sin duda ha de ser la lógica. El prototipo de inferencia racional es el silogismo «si P, entonces Q; P, por tanto Q». Consideremos un ejemplo sencillo.

Supongamos que las monedas de un país tienen un retrato de uno de sus eminentes soberanos en una cara y un retrato de uno de sus magníficos animales en la otra. Consideremos ahora una simple regla *si... entonces:* «Si una moneda tiene un rey en una cara, entonces tiene un ave en la otra». He aquí cuatro monedas que muestran un rey, una reina, un alce y un pato. ¿A cuáles tienes que darles la vuelta para determinar si se ha violado la regla?

Si eres como la mayoría de la gente, habrás dicho «el rey» o «el rey y el pato». La respuesta correcta es el rey y el alce. ¿Por qué? Todo el mundo está de acuerdo en que tenemos que dar la vuelta al rey, porque si no encontrásemos ningún ave en el reverso, se violaría expresamente la regla. La mayoría de la gente sabe que no tiene sentido dar la vuelta a la reina, porque la regla dice «si rey, entonces ave»; no dice nada acerca de las monedas con una reina. Muchos afirman que deberíamos dar la vuelta al pato, pero, bien pensado, esa moneda es irrelevante. La regla es: «Si rey, entonces ave», no «si ave, entonces rey»: si el pato compartiera la moneda con una reina, no habría ningún problema. Pero consideremos ahora el alce. Si diéramos la vuelta a la moneda y encontráramos un rey en el anverso, se habría transgredido la regla «si rey, entonces ave». Por lo tanto, la respuesta es el rey y el alce. Por término medio, solo el 10 % de las personas escogen esas monedas.

La tarea de selección de Wason (que lleva el nombre de su descubridor, el psicólogo cognitivo Peter Wason) se ha administrado desde hace sesenta y cinco años con varias reglas «si P, entonces Q» (la versión original utilizaba tarjetas con una letra en una cara y un número en la otra, y una regla como «si hay una D en una cara, hay un 3 en la otra»). Una y otra vez, la gente da la vuelta a P, o a P y Q, pero no a «NO Q».[24] No es que sean incapaces de entender la respuesta correcta. Al igual que en el test de reflexión cognitiva, tan pronto como se les explica, se dan una palmada en la frente y lo aceptan.[25] Ahora bien, abandonada a su suerte, su intuición irreflexiva no aplica la lógica.

¿Qué nos dice esto acerca de la racionalidad humana? Una explicación común es que revela nuestro *sesgo de confirmación*: el mal hábito de buscar pruebas que ratifiquen una creencia y mostrar indiferencia hacia las evidencias que podrían refutarla.[26] De este modo, los individuos piensan que los sueños son presagios porque recuerdan aquella vez en que soñaron

que un pariente tenía un percance y en efecto lo tuvo, pero olvidan todas las ocasiones en las que un pariente estaba bien y habían soñado que tenía un percance. O creen que los inmigrantes cometen muchos delitos porque leen una noticia acerca de un inmigrante que atracó una tienda, pero no piensan en el número mayor de tiendas atracadas por ciudadanos nacidos en el país.

El sesgo de confirmación es un diagnóstico común para la insensatez humana y un blanco para el fomento de la racionalidad. Francis Bacon (1561-1626), a quien se atribuye con frecuencia el desarrollo del método científico, escribió acerca de un hombre a quien llevaron a una iglesia y mostraron un cuadro de marineros que habían sobrevivido a un naufragio gracias a sus votos sagrados: «Sí —comentó—, pero ¿dónde están pintados aquellos que se ahogaron después de sus votos?».[27] Observó que «tal es la forma de todas las supersticiones, ya se trate de astrología, sueños, augurios, juicios divinos o cosas similares; los hombres, deleitándose en semejantes vanidades, reparan en los acontecimientos allí donde se cumplen, pero, allí donde no se cumplen, a pesar de que esto sucede mucho más a menudo, los ignoran y los pasan por alto».[28] Haciéndose eco de un célebre argumento del filósofo Karl Popper, la mayoría de los científicos actuales insisten en que la línea divisoria entre la ciencia y la pseudociencia es si los defensores de una hipótesis buscan deliberadamente pruebas capaces de falsarla y aceptan la hipótesis únicamente si esta sobrevive.[29]

¿Cómo pueden los humanos llegar al fin del día si son incapaces de aplicar la regla más elemental de la lógica? Parte de la respuesta es que la tarea de selección es un reto peculiar.[30] No pide a los individuos que apliquen el silogismo para hacer una deducción útil («he aquí una moneda con un rey; ¿qué hay en la otra cara?»), ni que comprueben la regla en general («¿es verdadera la regla para todo el sistema monetario del país?»). Les pregunta si la regla es aplicable específicamente a cada uno de los objetos que tienen delante de ellos sobre la mesa. La otra parte de la respuesta es que la gente sí que aplica la lógica cuando la regla implica los deberes y las prohibiciones de la vida humana en lugar de fichas y símbolos arbitrarios.

Supongamos que Correos vende sellos de cincuenta centavos para envíos de tercera clase (correos masivos por parte de empresas), pero requiere sellos de diez dólares para el correo urgente. Es decir, el correo propiamente dirigido debe seguir la regla: «Si una carta está etiquetada como correo urgente, ha de tener un sello de diez dólares». Supongamos que la etiqueta de la dirección y el sello no caben en la misma cara del sobre, de manera que el empleado de Correos tiene que dar la vuelta a cada sobre para comprobar si el remitente ha seguido la regla. He aquí cuatro sobres. Imagínate que eres un empleado de Correos. ¿A cuáles tienes que darles la vuelta?



Una vez más, la respuesta correcta es P y «NO Q», esto es, el sobre urgente y el que tiene un sello de cincuenta centavos. Aunque el problema es lógicamente equivalente al de las cuatro monedas, en esta ocasión casi todo el mundo acierta. El contenido de un problema lógico es relevante.[31] Cuando una regla *si... entonces* implementa un contrato que implica permisos y deberes («si disfrutas de un beneficio has de pagar un coste»), entonces una violación de la regla (beneficiarse sin pagar el coste) es equivalente a hacer trampa, y la gente sabe intuitivamente lo que hace falta para coger a un tramposo. No controlan a aquellos que no están disfrutando del beneficio, ni a quienes han pagado el coste, ninguno de los cuales podría estar intentando salirse con la suya.

Los psicólogos cognitivos debaten cuáles son exactamente las clases de contenidos que convierten a los individuos en lógicos de forma temporal. No puede tratarse solo de escenarios concretos, pero han de incorporar los tipos de desafíos lógicos con los que hemos llegado a sintonizar al convertirnos en adultos y quizá cuando evolucionamos hasta llegar a ser humanos. La supervisión de los privilegios o los deberes es uno de esos temas que desbloquean la lógica; otro de ellos es la supervisión de los peligros. La gente sabe que para verificar el cumplimiento de la precaución

«si montas en bicicleta debes llevar casco», tiene que controlar que un niño montado en una bicicleta lleve un casco y que un niño sin casco no monte en bicicleta.

Ahora bien, una mente capaz de falsar una regla condicional cuando las violaciones equivalen a un engaño o a un peligro no es exactamente una mente lógica. Por definición, la lógica se ocupa de la forma de los enunciados, no de su contenido: cómo P y Q están conectadas mediante *si, entonces, y, o, no, algunos* y *todos*, independientemente de a qué se refieran. La lógica es un logro supremo del conocimiento humano. Organiza nuestro razonamiento sobre temas desconocidos o abstractos, tales como las leyes del gobierno y de la ciencia, y, cuando se implementa en silicio, convierte la materia inerte en máquinas pensantes. Pero lo que controla la mente humana no instruida no es una herramienta de uso general y carente de contenido, con fórmulas como «[si P entonces Q] es equivalente a NO [P y NO Q]», en las que pueden introducirse cualesquiera P y Q. Controla un juego de herramientas más especializadas, que hornean los contenidos relevantes para el problema junto con las reglas de la lógica (sin dichas reglas, las herramientas no funcionarían). A la gente no le resulta fácil extraer las reglas y manejarlas en problemas nuevos, abstractos o aparentemente carentes de sentido. Para eso sirven la educación y otras instituciones que promueven la racionalidad. Estas aumentan la *racionalidad ecológica* con la que nacemos y crecemos —nuestro sentido común, nuestra astucia callejera— con herramientas de razonamiento de más amplio espectro y más potentes, perfeccionadas por nuestros mejores pensadores a lo largo de los milenios.[32]

UN SENCILLO PROBLEMA DE PROBABILIDAD

Uno de los más famosos concursos televisivos estadounidenses, en el apogeo del género entre los años cincuenta y ochenta del pasado siglo, fue *Let's Make a Deal [Hagamos un trato]*. Su presentador, Monty Hall, alcanzó una segunda clase de fama cuando se puso su nombre a un dilema de la teoría de la probabilidad, basado libremente en el programa.[33] Un concursante se enfrenta a tres puertas. Detrás de una de ellas hay un coche

nuevo y elegante. Tras las otras dos hay cabras. El concursante escoge una puerta, pongamos que la puerta 1. Para crear suspense, Monty abre una de las otras dos puertas, supongamos que la puerta 3, que revela una cabra. Para aumentar el suspense, ofrece al concursante la oportunidad de aferrarse a su elección original o cambiar a la puerta aún no abierta. Tú eres el concursante. ¿Qué deberías hacer?

Casi todo el mundo mantiene su opción.[34] Se figura que, dado que el coche se había colocado tras una de las tres puertas al azar y se ha eliminado la puerta 3, ahora existe una probabilidad de cincuenta-cincuenta de que el coche esté tras la puerta 1 o la puerta 2. Piensa que, aunque no pierde nada con cambiar, tampoco gana nada. Por consiguiente, mantiene su elección inicial por inercia, por orgullo o por la previsión de que su arrepentimiento tras un cambio desafortunado sería más intenso que su deleite tras uno afortunado.

El dilema de Monty Hall se hizo célebre en 1990 cuando se presentó en la columna «Ask Marilyn» [«Pregúntale a Marilyn»] de *Parade*, una revista inserta en la edición dominical de centenares de periódicos estadounidenses.[35] La columnista era Marilyn vos Savant, conocida a la sazón como «la mujer más inteligente del mundo» debido a su ingreso en el *Guinness Book of World Records (Libro Guinness de los récords mundiales)* por haber obtenido la puntuación más alta en el test de inteligencia. Vos Savant escribió que deberías cambiar: la probabilidad de que el coche esté detrás de la puerta 2 es de dos sobre tres, comparada con uno sobre tres para la puerta 1. La columna provocó diez mil cartas, un millar de ellas de doctores, principalmente en Matemáticas y Estadística, la mayoría de los cuales decían que estaba equivocada. He aquí algunos ejemplos:

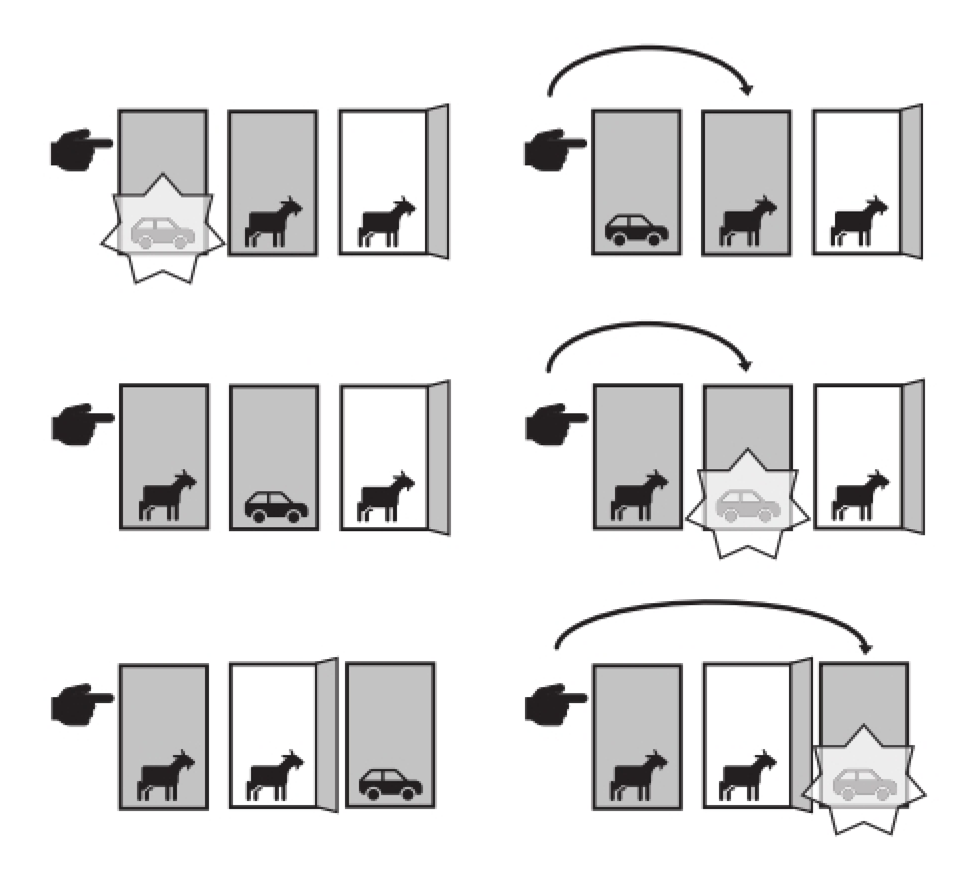
> ¡Ha metido usted la pata, y la ha metido hasta el fondo! Como parece tener dificultades para comprender el principio básico que está aquí en juego, se lo explicaré. Después de que el presentador revele una cabra, ya sabes que tienes una probabilidad de uno a dos de acertar. Tanto si cambias de opción como si no lo haces, las probabilidades son las mismas. Ya hay bastante analfabetismo matemático en este país y no necesitamos que lo propague más la persona con el cociente intelectual más alto del mundo. ¡Qué vergüenza!
>
> DOCTOR SCOTT SMITH, Universidad de Florida

> Estoy seguro de que recibirá muchas cartas sobre este asunto de alumnos de instituto y de universidad. Tal vez debería conservar unas cuantas direcciones para que la ayuden en sus futuras columnas. DOCTOR W. ROBERT SMITH, Universidad Estatal de Georgia Quizá las mujeres vean los problemas matemáticos de un modo diferente al de los hombres.
>
> DON EDWARDS, Sunriver, Oregón[36]

Entre los objetores estaba Paul Erdös (1913-1996), el renombrado matemático que fue tan prolífico que muchos académicos alardean de su «número de Erdös», la longitud de la cadena más corta de coautoría que los conecta con el gran teórico.[37]

No obstante, los matemáticos machistas y sabelotodo estaban equivocados y la mujer más inteligente del mundo tenía razón. Deberías cambiar. No cuesta tanto entender por qué. Existen tres posibles ubicaciones del coche. Consideremos cada puerta y contemos el número de veces que ganarías con cada estrategia. Escogiste la puerta 1, pero por supuesto esa es solo una etiqueta; siempre y cuando Monty siga la regla «abre una puerta no seleccionada con una cabra; si ambas tienen cabras, escoge una al azar», la probabilidad resulta ser la misma cualquiera que elijas.

Supongamos que tu estrategia es «mantenerte» (columna izquierda de la figura). Si el coche está detrás de la puerta 1 (arriba a la izquierda), ganas (no importa cuál de las otras dos puertas abra Monty, porque no cambias a ninguna de ambas). Si el coche está tras la puerta 2 (en el medio a la izquierda), pierdes. Si el coche está tras la puerta 3 (abajo a la izquierda), pierdes. Por tanto, la probabilidad de ganar con la estrategia de «mantenerte» es de uno sobre tres.



Supongamos ahora que tu estrategia es «cambiar» (columna derecha). Si el coche está tras la puerta 1, pierdes. Si el coche está tras la puerta 2, Monty habría abierto la puerta 3, de modo que tú cambiarías a la puerta 2 y ganarías. Si el coche está tras la puerta 3, él habría abierto la puerta 2, por lo que tú cambiarías a la puerta 3 y ganarías. Las probabilidades de ganar con la estrategia de «cambiar» son de dos sobre tres, el doble que la probabilidad de mantenerte.

No hace falta ser un genio.[38] Incluso si no examinas las posibilidades lógicas, podrías jugar tú mismo unas cuantas rondas con recortables y juguetes y sumar los resultados, como hizo el propio Hall para convencer a un escéptico periodista (hoy en día puedes jugar en línea).[39] O bien podrías seguir tu intuición: «Monty conoce la respuesta y me ha dado una pista; sería una estupidez no tomar nota de ella». ¿Por qué los matemáticos, profesores y demás peces gordos se equivocaron tan estrepitosamente?

Ciertamente hubo fracasos del pensamiento crítico derivados del sexismo, de los sesgos *ad hominem* y de la envidia profesional. Vos Savant es una mujer atractiva y elegante sin siglas después de su nombre, que escribía para una revistilla llena de recetas y cotilleos, y bromeaba en los programas nocturnos de entrevistas.[40] Desafiaba el estereotipo de una matemática, y su celebridad y su derecho a fanfarronear del *Guinness* la convertían en un blanco tremendamente atractivo.

Pero parte del problema radica en el problema mismo. Al igual que los señuelos en los test de reflexión cognitiva y de selección, el dilema de Monty Hall contiene algo diseñado para sacar a relucir la estupidez de nuestro sistema 1. Pero, en este caso, el sistema 2 no es mucho más brillante. Muchas personas son incapaces de digerir la explicación correcta incluso cuando esta se les señala. Una de ellas era Erdös, quien, violando el espíritu de un matemático, solo se convenció cuando vio el juego simulado reiteradamente.[41] Muchos persisten incluso cuando lo ven simulado y hasta cuando juegan repetidamente por dinero. ¿En qué consiste el desajuste entre nuestras intuiciones y las leyes de la probabilidad?

Una clave proviene de las justificaciones excesivamente confiadas que los sabiondos ofrecían para sus meteduras de pata, a veces transferidas irreflexivamente desde otros acertijos de la probabilidad. Mucha gente insiste en que cada una de las alternativas desconocidas (en este caso, las puertas sin abrir) ha de tener la misma probabilidad que las demás. Esto es cierto en el caso de los juguetes simétricos como las caras de una moneda o los lados de un dado, y es un punto de partida razonable cuando no sabemos absolutamente nada acerca de las alternativas. Ahora bien, no es una ley de la naturaleza.

Son muchos los que visualizan la cadena causal. El coche y las cabras estaban colocados antes de la revelación, y la apertura de una puerta no puede cambiarlos de sitio *a posteriori*. Señalar la independencia de los mecanismos causales es una manera habitual de echar por tierra otras ilusiones tales como la falacia del jugador, en la que la gente se equivoca al pensar que, después de una racha de rojos, el siguiente giro de la ruleta debería salir negro, cuando lo cierto es que la ruleta no tiene memoria, de manera que cada giro es independiente. Como explicaba con

condescendencia el autor de una de las cartas dirigidas a Vos Savant: «Imagínese una carrera con tres caballos, cada uno de los cuales tiene la misma probabilidad de ganar. Si el caballo 3 cae muerto a los quince metros de carrera, las probabilidades de cada uno de los dos caballos restantes ya no son de uno sobre tres, sino que pasan a ser de uno sobre dos». Claramente, concluía, no tendría sentido cambiar la apuesta del caballo 1 al caballo 2. Pero no es así como funciona el problema. Imagínate que, después de haber apostado por el 1, Dios anuncia: «No va a ser el caballo 3». Podría haber alertado contra el caballo 2, pero no lo hizo. Cambiar tu apuesta no parece tan disparatado.[42] En *Let's Make a Deal*, Monty Hall es Dios.

El divino presentador nos recuerda lo exótico que es el problema de Monty Hall. Requiere un ser omnisciente que desafíe el objetivo habitual de una conversación (compartir lo que tu oyente necesita saber —en este caso, qué puerta esconde el coche—) y que persiga, en cambio, la meta de aumentar el suspense de terceras personas.[43] Y a diferencia del mundo, cuyas pistas son indiferentes para nuestro trabajo detectivesco, el todopoderoso Monty sabe la verdad y conoce nuestra elección, y escoge su revelación en consecuencia.

La insensibilidad de la gente hacia esta lucrativa pero esotérica información identifica la debilidad que mora en el corazón del enigma: confundimos la *probabilidad* con la *propensión*. Una propensión es la disposición de un objeto a actuar de ciertas maneras. Las intuiciones acerca de las propensiones constituyen una parte fundamental de nuestros modelos mentales del mundo. La gente siente que las ramas dobladas tienden a volver a su forma original, que los kudús pueden cansarse fácilmente, que los puercoespines dejan con frecuencia rastros con dos huellas de almohadillas. Una propensión no puede percibirse directamente (o bien la rama se dobló o bien no lo hizo), pero puede inferirse escudriñando la constitución física de un objeto y trabajando con las leyes de causa y efecto. Una rama más seca puede romperse, un kudú tiene más resistencia en la temporada de lluvias, un puercoespín tiene dos almohadillas proximales que dejan un rastro de almohadillas cuando el suelo es blando, pero no necesariamente cuando es duro.

Pero la probabilidad es diferente; es una herramienta conceptual inventada en el siglo XVII.[44] La palabra posee varios significados, pero el relevante a la hora de tomar decisiones arriesgadas es la fortaleza de nuestra creencia en un estado de cosas desconocido. Cualquier pizca de evidencia que altere nuestra confianza en un resultado cambiará su probabilidad, así como la manera racional de actuar en consecuencia. La dependencia de la probabilidad respecto de conocimientos etéreos más que de la mera constitución física contribuye a explicar por qué la gente falla en el dilema. Intuye las propensiones del coche a terminar detrás de las diferentes puertas y sabe que la apertura de una puerta no podría haber modificado esas propensiones. Pero las probabilidades no tienen que ver con el mundo; tienen que ver con nuestra *ignorancia* acerca del mundo. Las nuevas informaciones reducen nuestra ignorancia y alteran la probabilidad. Si esto suena místico o paradójico, piensa en la probabilidad de que una moneda que acabo de lanzar al aire haya caído de cara. Para ti es 0,5. Para mí es 1 (he mirado a hurtadillas). El mismo suceso, diferente conocimiento, distinta probabilidad. En el dilema de Monty Hall, el presentador que todo lo ve ofrece nueva información.

Una consecuencia es que, cuando la reducción de la ignorancia otorgada por el presentador está vinculada con más transparencia de las circunstancias físicas, la solución al problema deviene intuitiva. Vos Savant invitaba a sus lectores a imaginar una variante del concurso con, pongamos por caso, mil puertas.[45] Escoges una. Monty revela una cabra tras 998 de las restantes. ¿Cambiarías a la puerta que ha dejado cerrada? En esta ocasión parece evidente que la elección de Monty transmite información procesable. Podemos imaginarle escaneando las puertas en busca del coche mientras decide cuál no abrir, y la puerta cerrada es una señal de que ha localizado el coche y, por ende, un rastro del propio coche.

UN SENCILLO PROBLEMA DE PREDICCIÓN

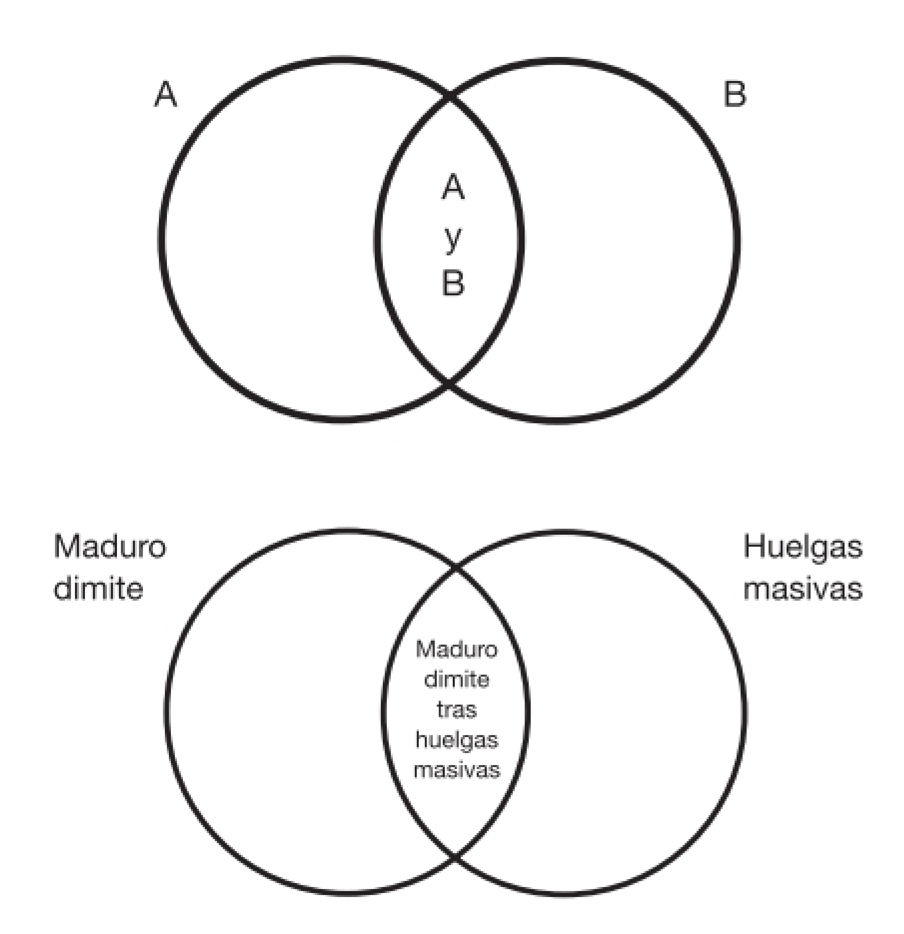
Una vez que nos habituamos a asignar números a los sucesos desconocidos, podemos cuantificar nuestras intuiciones acerca del futuro. La predicción de sucesos es una empresa de envergadura. Sirve de

fundamento para las políticas, las inversiones, la gestión de riesgos y la curiosidad ordinaria acerca de lo que le aguarda al mundo. Considera cada uno de los sucesos siguientes y anota tus estimaciones de la probabilidad de que tenga lugar en la próxima década. Muchos de ellos son bastante improbables, de modo que hagamos distinciones más finas en el extremo inferior de la escala y escojamos una de las siguientes probabilidades para cada uno de ellos: menos de 0,01 %, 0,1 %, 0,5 %, 1 %, 2 %, 5 %, 10 %, 25 % y 50 % o más.

1. Arabia Saudí desarrolla un arma nuclear.
2. Nicolás Maduro dimite como presidente de Venezuela.
3. Rusia tiene una presidenta.
4. El mundo sufre una nueva pandemia, más letal aún que la COVID-19.
5. Se impide constitucionalmente a Vladimir Putin que se presente a un nuevo mandato como presidente de Rusia y su mujer ocupa su lugar en los comicios, permitiéndole gobernar el país desde la barrera.
6. Huelgas y disturbios masivos fuerzan a Nicolás Maduro a dimitir como presidente de Venezuela.
7. Un virus respiratorio salta de los murciélagos a los humanos en China y comienza una nueva pandemia, más letal aún que la COVID-19.
8. Después de que Irán desarrolle un arma nuclear y la pruebe en una explosión subterránea, Arabia Saudí desarrolla un arma nuclear en respuesta.

Presenté ítems como estos a varios centenares de participantes en una encuesta. Por término medio, la gente consideraba más probable que la mujer de Putin fuese presidenta de Rusia que que hubiese una presidenta. Consideraban más probable que las huelgas forzasen a Maduro a dimitir que que dimitiese. Consideraban más probable que Arabia Saudí desarrollase un arma nuclear en respuesta a una bomba iraní que que llegase a desarrollar un arma nuclear. Y consideraban más probable que los murciélagos chinos comenzasen una pandemia que que hubiese una pandemia.[46]

Probablemente estés de acuerdo al menos con una de estas comparaciones, al igual que el 86 % de los participantes que valoraron todos los ítems. En tal caso, has violado una ley elemental de la probabilidad, la regla de conjunción: la probabilidad de una conjunción de sucesos (A Y B) ha de ser igual o menor que la probabilidad de cualquiera de los sucesos (A O B). La probabilidad de coger una pica con un número par en una baraja, por ejemplo (par y pica), tiene que ser menor que la probabilidad de coger una pica, porque algunas picas no son números pares.



En cada par de sucesos del mundo, el segundo escenario es una conjunción de sucesos, uno de los cuales es el suceso del primer escenario. Por ejemplo, «Irán prueba un arma nuclear y Arabia Saudí desarrolla un arma nuclear» es una conjunción que abarca «Arabia Saudí desarrolla un arma nuclear» y debe tener una probabilidad menor de suceder, ya que hay otros escenarios en los que los saudíes podrían desarrollar armas nucleares

(para hacer frente a Israel, para hacer alarde de hegemonía en el golfo Pérsico, etcétera). Siguiendo la misma lógica, la dimisión de Maduro de la presidencia de Venezuela tiene que ser más probable que la dimisión de Maduro de la presidencia tras una serie de huelgas.

¿Qué piensa la gente? Una clase de acontecimientos descritos mediante un único enunciado puede ser genérica y abstracta, sin nada a lo que la mente pueda aferrarse. Una clase de acontecimientos descritos mediante una conjunción de enunciados puede ser más vívida, especialmente cuando estos detallan una trama que somos capaces de ver en el teatro de nuestra imaginación. La probabilidad intuitiva está impulsada por la imaginabilidad: cuanto más fácil de visualizar es algo, más probable parece. Esto nos hace incurrir en lo que Tversky y Kahneman denominan la *falacia de la conjunción*, en virtud de la cual una conjunción es intuitivamente más probable que cualquiera de los elementos que la integran.

Los pronósticos de los expertos están impulsados con frecuencia por vívidas narrativas que dan al traste con la probabilidad.[47] Un famoso artículo publicado en la portada de *The Atlantic* en 1994 por el periodista Robert Kaplan predecía «la anarquía que viene».[48] Kaplan pronosticaba que en las primeras décadas del siglo XXI se librarían guerras por recursos escasos como el agua; Nigeria conquistaría Níger, Benín y Camerún; habría guerras mundiales por el control de África; Estados Unidos, Canadá, la India, China y Nigeria se desintegrarían, a raíz de lo cual los hispanoestadounidenses querrían unirse a México y Alberta desearía unirse a Montana; aumentaría la delincuencia en las ciudades estadounidenses; el sida iría de mal en peor; amén de otra docena de calamidades, crisis y colapsos. Sin embargo, mientras el artículo causaba sensación (incluso al presidente Bill Clinton, que lo hizo circular por la Casa Blanca), el número de guerras civiles, la proporción de personas sin acceso al agua salubre y la tasa de delincuencia en Estados Unidos caían en picado.[49] A los tres años, un efectivo tratamiento para el sida comenzaría a diezmar el número de muertos por esta enfermedad. Y más de un cuarto de siglo después, las fronteras nacionales apenas se han movido.

La falacia de la conjunción fue ilustrada por vez primera por Tversky y Kahneman con un ejemplo que se ha hecho célebre como «el problema de Linda».[50]

> Linda tiene treinta y un años, es soltera, no tiene pelos en la lengua y es muy brillante. Se graduó en Filosofía. En sus años de estudiante, estaba profundamente preocupada por los problemas de discriminación y justicia social, y participaba asimismo en manifestaciones antinucleares.
>
> Por favor, indique la probabilidad de cada uno de estos enunciados:
>
> - Linda es profesora de escuela primaria.
> - Linda es activista del movimiento feminista.
> - Linda es trabajadora social psiquiátrica.
> - Linda es cajera de un banco.
> - Linda es vendedora de seguros.
> - Linda es cajera de un banco y activista del movimiento feminista.

Los encuestados juzgaban más probable que Linda fuese una cajera de banco feminista que que fuese cajera de banco: una vez más, se consideraba que la probabilidad de A Y B era más alta que la probabilidad de solo A. La anticuada estampa —con su «Linda» de la generación del *baby boom*, el ambiguo cumplido de ser *brillante*, las protestas *demodés* y su ocupación en declive— exhibe el color sepia de principios de los años ochenta. Pero, como cualquier profesor de Psicología sabe, el efecto es fácilmente replicable y, hoy en día, la inteligentísima Amanda que participa en las marchas de Black Lives Matter sigue teniendo más probabilidades de ser una enfermera titulada feminista que una enfermera titulada.

El problema de Linda activa nuestras intuiciones de una manera particularmente convincente. A diferencia de acertijos como la tarea de selección, en la que las personas cometen errores cuando el problema es abstracto («si P, entonces Q») y aciertan cuando se formula en ciertos escenarios de la vida real, aquí todo el mundo está de acuerdo con la ley abstracta de que $P(A \cap B) \leq P(A)$ —la probabilidad de A y a la vez B es menor o igual que la probabilidad de A—, pero cambia drásticamente de parecer cuando esta se concreta. El biólogo y divulgador científico Stephen Jay Gould hablaba en nombre de muchos cuando comentaba: «Ya sé que el

enunciado [conjuntivo] es menos probable, pero un pequeño homúnculo no deja de brincar en mi cabeza gritándome: “Pero no puede ser solo una cajera de banco; lee la descripción”».[51]

Ese pequeño homúnculo puede ser explotado por los persuasores diestros. Un fiscal con poco con lo que trabajar, aparte del cadáver de una mujer hallado en una playa, puede inventarse una historia sobre cómo su marido podría, hipotéticamente, haberla asfixiado y haber arrojado el cuerpo para poder casarse con su amante e iniciar un negocio con el dinero del seguro. El abogado defensor podría contar una historia alternativa en la que, en teoría, la mujer podría haber sido la víctima de un robo de bolso nocturno con un desenlace fatal. Según las leyes de la probabilidad, cada detalle conjetural debería tornar menos probable el escenario y, sin embargo, cada uno puede volverlo más convincente. Como decía Poo-Bah en *The Mikado (El mikado)*, se trata de «detalles meramente corroborativos, que pretenden conferir verosimilitud artística a una narración por lo demás escueta y poco convincente».[52]

La regla de la conjunción es una ley básica de la probabilidad matemática, y no necesitamos pensar en números para entenderla. Esto hizo que Tversky y Kahneman fuesen pesimistas en lo tocante al sentido de probabilidad intuitivo de los individuos que, a su juicio, está impulsado por los estereotipos representativos y los recuerdos disponibles, en lugar de por un cálculo sistemático de posibilidades. Rechazaban la idea de que «dentro de cada persona incoherente hay otra coherente que intenta salir».[53]

Otros psicólogos son más caritativos. Como vimos con el dilema de Monty Hall, el término «probabilidad» *(probability)* posee varios significados, entre los que se incluyen la propensión física, la fuerza justificada de la creencia y la frecuencia a largo plazo. Otro sentido aparece recogido en el *Oxford English Dictionary [Diccionario de inglés de Oxford]*: «La apariencia de verdad o verosimilitud *(likelihood)* que comporta cualquier enunciado o suceso a la luz de las evidencias actuales».[54] Quienes se enfrentan al problema de Linda saben que «la frecuencia a largo plazo» es irrelevante: solamente hay una Linda y, o bien es una cajera de banco feminista, o bien no lo es. En cualquier conversación coherente, el hablante ofrecería detalles biográficos por una razón: conducir

al oyente a una conclusión plausible. Según los psicólogos Ralph Hertwig y Gerd Gigerenzer, los individuos pueden haber inferido racionalmente que el significado relevante de *probabilidad* en esta tarea no es uno de los sentidos matemáticos en los que resulta aplicable la regla de la conjunción, sino el sentido no matemático de «grado de justificación a la luz de las evidencias actuales», y pueden haber seguido sensatamente lo apuntado por las pruebas.[55]

En respaldo de la interpretación caritativa, muchos estudios, comenzando con uno de los propios Tversky y Kahneman, muestran que, cuando *se anima* a las personas a razonar acerca de la probabilidad en el sentido de la frecuencia relativa, en lugar de dejar que se enfrenten al enigmático concepto de *probabilidad de un caso único*, es más probable que obedezcan la regla de la conjunción. Imagínate a mil mujeres como Linda. ¿Cuántas crees que son cajeras de banco? ¿Cuántas crees que son cajeras de banco a la par que activistas del movimiento feminista? Ahora el homúnculo guarda silencio; una persona coherente intenta salir. La tasa de errores de la conjunción cae en picado.[56] ¿Es entonces la falacia de la conjunción, la demostración por antonomasia de la ceguera humana a la probabilidad, un artefacto de la redacción ambigua y las preguntas capciosas? Tversky y Kahneman insistían en que no lo es. Observaban que las personas cometen la falacia incluso cuando se las invita a *apostar* por las posibilidades (sí, una mayoría prefiere apostar a que Linda es una cajera feminista antes que a que solo es una cajera). E incluso cuando la pregunta se formula en frecuencias, de modo que los individuos podrían evitar un error de la conjunción simplemente contando las cajeras de banco en su imaginación, una minoría sustancial lo comete. Esta se convierte en mayoría cuando los individuos consideran cada alternativa de forma aislada en lugar de ver una al lado de la otra, y no les restriega por la cara la absurdidad de que un subconjunto supere en número a un superconjunto.[57]

Kahneman ha observado que los humanos nunca son tan irracionales como cuando protegen sus ideas favoritas. Por consiguiente, propugnaba un nuevo método para resolver las controversias científicas con el fin de reemplazar la inveterada tradición de que los rivales se turnen para cambiar las reglas de juego y descalificarse mutuamente en una salva de réplicas. En

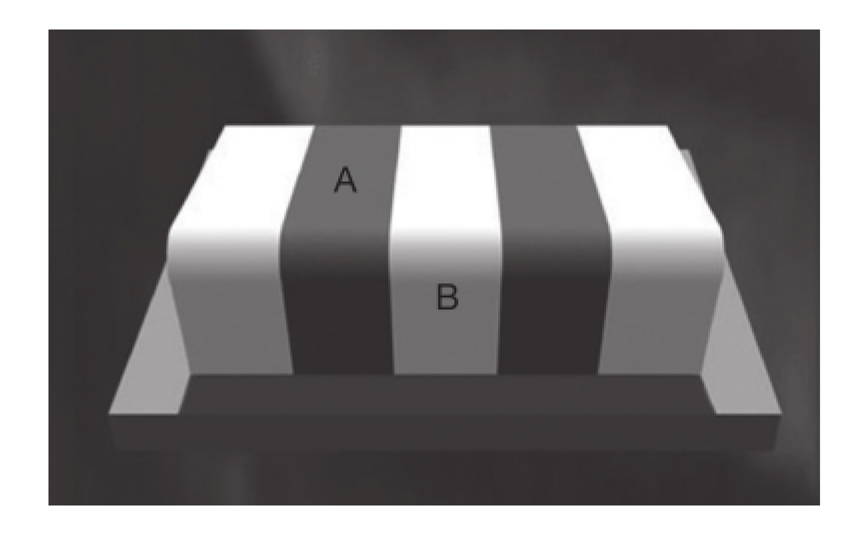
una «colaboración contradictoria» *(adversarial collaboration)*, los contendientes acuerdan de antemano una prueba empírica que resolvería la cuestión e invitan a un árbitro a unirse a ellos para llevarla a cabo.[58] Convenientemente, Kahneman colaboró con Hertwig para ver quién estaba en lo cierto respecto al problema de Linda, reclutando a la psicóloga Barbara Mellers para que actuase como árbitra. El equipo de rivales acordó realizar tres estudios que formulasen el problema en frecuencias («de cien personas como Linda, ¿cuántas son...?», en lugar de preguntar solo por Linda. En su informe sobre los complejos resultados, el trío explicaba: «No pensábamos que los experimentos fuesen a resolver todos los problemas ni se produjo ningún milagro». Pero ambas partes convenían en que las personas son propensas a cometer la falacia de la conjunción incluso cuando tratan con frecuencias. Y coincidían en que, bajo las circunstancias adecuadas (las alternativas se encuentran disponibles por comparación una al lado de otra, y la formulación de las alternativas no deja nada a la imaginación), la gente puede hallar la manera de no incurrir en dicha falacia.

LA MORALEJA DE LAS ILUSIONES COGNITIVAS

¿Cómo reconciliamos la racionalidad que permite a nuestra especie sobrevivir con su ingenio en entornos antiguos y modernos con las meteduras de pata reveladas por estos rompecabezas: el sesgo de confirmación, el exceso de confianza, la distractibilidad en los detalles concretos y los hábitos conversacionales? Los errores clásicos del razonamiento se denominan con frecuencia *ilusiones cognitivas* y los paralelismos con las ilusiones visuales familiares por las cajas de cereales y los museos de ciencias resultan instructivos. Son más profundos que el hecho evidente de que nuestros ojos y nuestra mente pueden engañarnos. Explican cómo nuestra especie puede ser tan inteligente y, sin embargo, tan fácil de engañar.

He aquí dos ilusiones clásicas, que cobran vida gracias al neurocientífico Beau Lotto.[59] La primera es una ilusión del sombreado. Aunque parezca mentira, las franjas oscuras de la parte superior de la caja

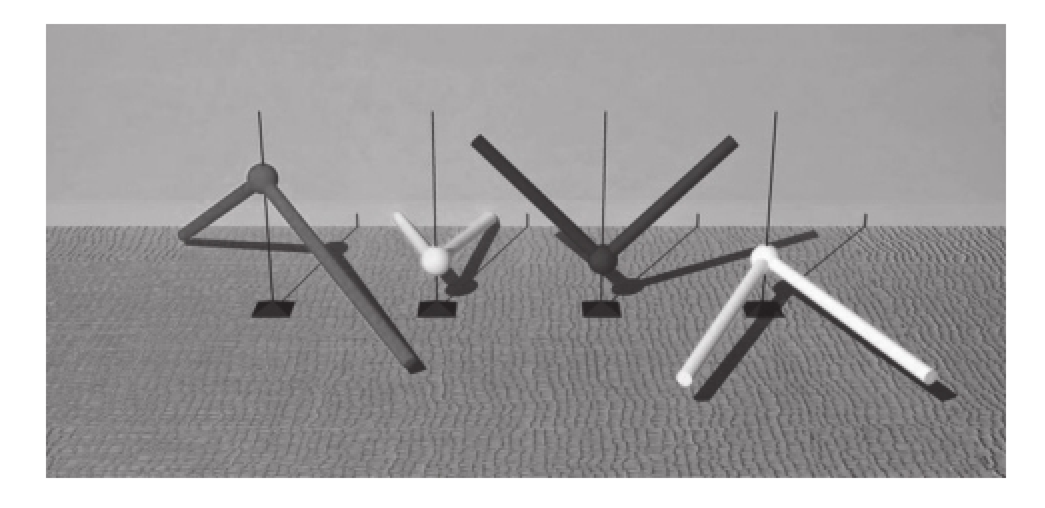
(A) y las franjas blancas de la parte delantera (B) son tonos idénticos de gris.



Utilizado con el permiso de Beau Lotto.

La segunda es una ilusión de la forma: los ángulos de los cuatro codos son idénticos: noventa grados.

La primera lección es que no siempre podemos fiarnos de nuestros ojos o, para ser más precisos, del sistema visual 1 de nuestro cerebro. La segunda es que podemos reconocer nuestros errores utilizando el sistema 2, pongamos por caso, cubriendo la primera figura con una tarjeta con dos perforaciones y alineando el ángulo de la tarjeta con los codos en la segunda.



Utilizado con el permiso de Beau Lotto.

Pero la moraleja equivocada es que el sistema visual humano es un defectuoso artilugio que no cesa de engañarnos con figuraciones y espejismos. El sistema visual humano es una de las maravillas del mundo. Es un instrumento de precisión capaz de detectar un solo fotón, reconocer miles de formas y tomar senderos pedregosos y autopistas de alta velocidad. Supera nuestros mejores sistemas de visión artificial, y ese es el motivo por el cual, en el momento de escribir estas páginas, todavía no se han soltado por las calles de las ciudades los vehículos autónomos, pese a las décadas de I+D. Los módulos de visión de los coches robots tienden a confundir un semirremolque con una valla publicitaria o una señal de tráfico cubierta de pegatinas con un frigorífico lleno de comida.[60]

Las ilusiones de la forma y del sombreado no son errores, sino características. El objetivo del sistema visual es proporcionar al resto del cerebro una descripción precisa de las formas tridimensionales y la composición material de los objetos que tenemos delante de nosotros.[61] Este es un problema arduo, porque la información que llega al cerebro desde la retina no refleja directamente la realidad. La luminosidad de una mancha en la imagen retiniana no depende exclusivamente de la pigmentación de la superficie en el mundo, sino también de la intensidad de la iluminación que cae sobre ella: una mancha gris podría haber surgido de una superficie negra iluminada por una luz intensa o de una superficie blanca iluminada por una luz tenue (esta es la base de la ilusión llamada #thedress [el vestido], que arrasó en internet en 2015).[62] Una forma en la retina no depende solo de la geometría tridimensional del objeto, sino también de su orientación desde un punto de vista determinado: un ángulo agudo en la retina podría ser una esquina aguda vista de frente o una esquina en ángulo recto en escorzo. El sistema visual anula los efectos de estas distorsiones, dividiendo la intensidad de la iluminación e invirtiendo la trigonometría de la perspectiva a fin de suministrar al resto del cerebro una representación que se corresponda con las formas y los materiales del mundo. El bloc de notas intermedio en estos cálculos (el despliegue bidimensional de píxeles procedentes de nuestra retina) se oculta a los sistemas de razonamiento y planificación del cerebro, pues no aportaría más que distracciones.

Gracias a este diseño, nuestro cerebro no es muy bueno como fotómetro ni como transportador de ángulos, pero tampoco necesita serlo (a menos que seamos pintores realistas). Las ilusiones surgen cuando se pide a las personas que funcionen precisamente como esos instrumentos. Se pide al observador que repare en la luminosidad de la franja o en la medida del ángulo *en la imagen*. Las imágenes se han confeccionado de tal manera que las propiedades simples (la misma luminosidad, los ángulos rectos) están enterradas en el bloc de notas que la mente consciente habitualmente ignora. Si las preguntas se refiriesen a las cosas del *mundo* captadas en las imágenes, nuestras impresiones serían correctas. La franja gris es realmente más oscura que la franja blanca tanto en la cara iluminada de la caja como en la sombreada; los codos dispuestos con diferentes inclinaciones están realmente doblados en diferentes ángulos.

Análogamente, las ilusiones cognitivas como las de este capítulo pueden surgir porque dejamos de lado el planteamiento literal de una pregunta tal como esta llega a nuestro cerebro y pensamos detenidamente en lo que preguntaría de manera razonable un hablante en el mundo social. Hacer aritmética con números engañosamente conspicuos, verificar una proposición acerca de un puñado de fichas, escoger a partir de las pistas ofrecidas por un maestro astuto y omnisciente, y seguir un vívido bosquejo de un personaje hasta una conclusión literal, pero implausible, son tareas un tanto semejantes a juzgar los ángulos y los tonos de gris en la página impresa. Nos conducen a respuestas incorrectas, sí, pero se trata con frecuencia de respuestas correctas a preguntas diferentes y más útiles. Una mente capaz de interpretar el propósito de un interrogador en su contexto es sumamente sofisticada. Por eso pulsamos furiosamente el «0» y gritamos «operador» cuando el bot de una línea de asistencia telefónica repite una lista de opciones inútiles y solo un humano puede llegar a comprender el motivo de nuestra llamada.

El hecho de que nuestras reacciones irracionales resulten explicables no es una excusa para recurrir a ellas, en mayor medida de lo que deberíamos fiarnos siempre de nuestros ojos. La ciencia y la tecnología han ampliado de forma emocionante las capacidades del sistema visual más allá de lo que la naturaleza nos ha otorgado. Disponemos de microscopios para

lo pequeño, telescopios para lo distante, fotografías para lo pasado, iluminación para lo oscuro, teledetección para lo invisible. Y, cuando nos adentramos en territorios que están fuera del entorno en el que hemos evolucionado, como lo muy rápido y lo muy alto, confiar en nuestros sentidos puede ser fatal. Los juicios de profundidad y orientación que permiten a nuestro cerebro anular los efectos de la geometría proyectiva en la vida cotidiana dependen de las líneas convergentes, las texturas en retroceso y los contornos fluidos dispuestos a lo largo del terreno mientras nos movemos y miramos a nuestro alrededor. Cuando un piloto se halla suspendido a miles de metros en el aire sin nada más que el espacio vacío entre él y la tierra, y el horizonte está oscurecido por las nubes, la bruma o las montañas, su sentido visual pierde la sincronía con la realidad. Mientras vuela guiándose por su intuición, sin ser capaz de distinguir la aceleración de la gravedad, cualquier corrección no hace sino empeorar las cosas y puede enviar el avión a una «espiral de cementerio» en cuestión de minutos, el destino de un inexperto y excesivamente confiado John F. Kennedy Jr. en 1999. Por muy excelentes que sean nuestros sistemas visuales, los pilotos racionales saben cuándo descartarlos y dirigir su percepción hacia los instrumentos.[63]

Y, por muy excelentes que sean nuestros sistemas cognitivos, en el mundo moderno hemos de saber cuándo dejarlos de lado y dirigir nuestro razonamiento hacia los instrumentos: las herramientas de la lógica, la probabilidad y el pensamiento crítico, que amplían nuestras capacidades de razonar más allá de las que la naturaleza nos ha otorgado. Porque en el siglo XXI, cuando pensamos guiándonos por nuestra intuición, cualquier corrección puede empeorar las cosas y puede enviar nuestra democracia a una espiral de cementerio.

CAPÍTULO

2

Racionalidad e irracionalidad

¿Podría decir que no disfruto en absoluto sirviendo con humanos? Sus ilógicas y estúpidas emociones no dejan de irritarme.

MR. SPOCK

La racionalidad está fuera de onda. Describir a alguien con palabras de la jerga inglesa como *nerd*, *wonk*, *geek* o *brainiac* para referirse a las personas cerebrales supone dar a entender que está irremediablemente desfasado. Durante décadas, los guiones de Hollywood y las letras de las canciones de *rock* han equiparado la alegría y la libertad con escapar de la razón. «Un hombre necesita un poco de locura; si no, nunca se atreverá a cortar la soga y ser libre», decía Zorba el Griego. «Deja de tener sentido», aconsejaban Talking Heads; «hagamos locuras», exhortaba el artista anteriormente conocido como Prince.* Movimientos académicos de moda como el posmodernismo y la teoría crítica (que no hemos de confundir con el pensamiento crítico) sostienen que la razón, la verdad y la objetividad son construcciones sociales que justifican el privilegio de los grupos dominantes. Estos movimientos tienen un aire de sofisticación, que implica que la filosofía y la ciencia occidentales son provincianas, anticuadas e ingenuas ante la diversidad de formas de conocer halladas a lo largo de las épocas y las culturas. Es cierto que no lejos de donde vivo, en el centro de Boston, hay un espléndido mosaico turquesa y dorado que proclama: «Sigue a la razón». Pero está colocado en la Gran Logia de los Masones, la organización fraternal del fez y el mandil que es la respuesta a la pregunta: «¿Qué es lo contrario de estar en la onda?».

Mi posición personal con respecto a la racionalidad es que «estoy a favor de ella». Aunque no sea capaz de argüir que la razón es guay, súper, genial, la bomba, la leche o el no va más, y, en rigor, ni siquiera pueda justificar ni racionalizar la razón, defenderé el mensaje del mosaico: deberíamos *seguir* a la razón.

RAZONES A FAVOR DE LA RAZÓN

Empecemos por el principio: ¿qué es la racionalidad? Como sucede con la mayoría de las palabras de uso común, ninguna definición puede estipular su significado con exactitud y el diccionario solo nos conduce a un

círculo: la mayoría define *racional* como «dotado de razón», pero la *razón* misma proviene del latín *ratio-*, definida con frecuencia como «razón».

Una definición más o menos fiel a la forma en que se usa la palabra es «la capacidad de utilizar el conocimiento para alcanzar objetivos». El *conocimiento* suele definirse, a su vez, como «creencia verdadera y justificada».[1] No atribuiríamos a alguien la condición de racional si actuara basándose en creencias que se supiese que son falsas, como buscar sus llaves en un lugar en el que supiera que no podían estar, o bien si esas creencias no pudieran estar justificadas, por ejemplo, si fuesen fruto de una visión inducida por las drogas o por una voz alucinada, en lugar de resultar de la observación del mundo o de la inferencia a partir de alguna creencia verdadera.

Además, las creencias han de profesarse al servicio de un objetivo. A nadie se le atribuye la racionalidad meramente por pensar cosas verdaderas, como calcular los dígitos de π o producir las implicaciones lógicas de una proposición («o 1 + 1 = 2 o la luna está hecha de queso», «si 1 + 1 = 3, entonces los cerdos vuelan»). Un agente racional debe tener un *objetivo*, ya sea este determinar la verdad de una idea digna de ser tenida en cuenta, lo cual se conoce como razón teórica, ya sea lograr un resultado notable en el mundo, lo cual se conoce como razón práctica («qué es verdad» y «qué hacer»). Incluso la racionalidad rutinaria de ver en lugar de alucinar está al servicio del objetivo siempre presente e inherente a nuestros sistemas visuales de conocer nuestro entorno.

Por otra parte, un agente racional ha de alcanzar dicho objetivo no haciendo algo que simplemente resulta funcionar en ese momento y en ese lugar, sino empleando algún conocimiento relevante para las circunstancias. Así es como distinguía William James una entidad racional de otra no racional que, en un principio, parecería estar haciendo lo mismo:

> Romeo quiere a Julieta como las limaduras al imán; y si no interviene ningún obstáculo, él avanza hacia ella en línea recta, tal como las limaduras. Pero Romeo y Julieta, si se levantara un muro entre ellos, no permanecerían tontamente oprimiéndose la cara contra sus lados opuestos, como el imán y las limaduras [...]. Romeo pronto encuentra un camino indirecto de besar directamente los labios de Julieta, escalando el muro o de otra forma. Con las limaduras el camino es fijo; el que lleguen al final depende de accidentes. Con el amante, el final es lo que es fijo; el camino puede modificarse indefinidamente.[2]

Con esta definición, la defensa de la racionalidad parece demasiado evidente: ¿quieres las cosas o no las quieres? Si las quieres, la racionalidad es lo que te permite conseguirlas.

Ahora bien, esta defensa de la racionalidad está expuesta a una objeción. Nos aconseja que fundamentemos nuestras creencias sobre la verdad, con el fin de garantizar que nuestra inferencia desde una creencia hasta otra esté justificada, y que hagamos planes que tengan probabilidades de conseguir un determinado fin. Pero eso no hace sino plantear interrogantes como qué es la *verdad*, qué hace que una inferencia esté «justificada», o cómo sabemos que podemos hallar medios que logren realmente un fin determinado. Pero el afán de ofrecer la razón última, absoluta y definitiva para la razón es una empresa descabellada. Al igual que un niño inquisitivo de tres años responderá a cada respuesta a un porqué con otro porqué, el intento de hallar la razón última para la razón siempre puede ser obstaculizado por la demanda de ofrecer una razón para la razón de la razón. Solo por creer que P implica Q y creer P, ¿por qué debería creer Q? ¿Es porque también creo que [(P implica Q) y P] implica Q? Pero ¿por qué debería creer *eso*? ¿Acaso porque tengo aún otra creencia: {[(P implica Q) y P] implica Q} implica Q?

Esta regresión era la base de la historia de Lewis Carroll de 1895 «What the tortoise said to Achilles» («Lo que la tortuga le dijo a Aquiles»), que imaginaba la conversación que tendría lugar cuando el guerrero de los pies ligeros alcanzó (pero jamás fue capaz de adelantar) a la tortuga que le llevaba ventaja en la segunda paradoja de Zenón (en el tiempo que tardaba Aquiles en colmar la brecha, la tortuga avanzaba, abriendo una nueva brecha que Aquiles debía colmar, *ad infinitum*). Además de escritor infantil, Carroll era un lógico y, en este artículo publicado en la revista de filosofía *Mind*, se imagina al guerrero sentado a lomos de la tortuga y respondiendo a las crecientes demandas de esta para justificar sus argumentos llenando un cuaderno con miles de reglas para las reglas de las reglas.[3] La moraleja es que el razonamiento con reglas lógicas en algún punto ha de *ejecutarse* simplemente mediante un mecanismo incorporado a la máquina o innato en el cerebro, que actúa porque así es como funcionan los circuitos, no porque consulte una regla que le diga lo que ha de hacer. Programamos

aplicaciones en un ordenador, pero la CPU de este no es ella misma una *app*; es un trozo de silicio en el que se han grabado operaciones elementales, como comparar símbolos y añadir números. Estas operaciones están diseñadas (por un ingeniero o, en el caso del cerebro, por la selección natural) para implementar leyes de la lógica y de las matemáticas que son inherentes al ámbito abstracto de las ideas.[4]

Ahora bien, a pesar del señor Spock, la lógica no es lo mismo que el razonamiento, y en el próximo capítulo exploraremos las diferencias. No obstante, ambos están íntimamente relacionados, y las razones por las que las reglas de la lógica no pueden ser ejecutadas mediante más reglas de la lógica todavía *(ad infinitum)* son aplicables asimismo a la justificación de la razón mediante más razón aún. En cada caso, la regla definitiva ha de ser: «Simplemente hazlo». Al fin y a la postre, los participantes en una discusión no tienen más opción que comprometerse a razonar, porque eso es lo que se comprometieron a hacer al comienzo, cuando entablaron una conversación sobre por qué deberíamos seguir la razón. Siempre y cuando las personas estén argumentando y persuadiendo, y luego evaluando y aceptando o rechazando los argumentos —en vez de, pongamos por caso, dedicarse a sobornarse o a amenazarse mutuamente para pronunciar ciertas palabras—, ya es demasiado tarde para preguntar por el valor de la razón. Ya están razonando y han aceptado tácitamente el valor de esta.

Cuando se trata de argumentar en contra de la razón, tan pronto como apareces, pierdes. Supongamos que aduces que la racionalidad es innecesaria. ¿Es racional *ese* enunciado? Si concedes que no lo es, entonces no hay razón para que yo lo crea, como tú mismo acabas de decir. Pero si insistes en que debo creerlo porque el enunciado es racionalmente convincente, has reconocido que la racionalidad es la medida mediante la cual deberíamos aceptar las creencias, en cuyo caso esa en particular ha de ser falsa. Análogamente, si afirmaras que todo es subjetivo, yo podría preguntar: «¿Es subjetivo *ese* enunciado?». Si lo es, entonces tú eres libre de creerlo, pero yo no tengo por qué hacerlo. O supón que sostienes que todo es relativo. ¿Es relativo *ese* enunciado? Si lo es, entonces puede ser verdadero para ti aquí y ahora, pero no para cualquier otro o cuando hayas dejado de hablar. Por eso no puede ser cierto el reciente cliché de que

vivimos en una «era de la posverdad». Si fuera verdadero, entonces no sería verdadero, porque ello supondría afirmar algo verdadero acerca de la época en la que vivimos.

Este argumento, expuesto por el filósofo Thomas Nagel en *The Last Word (La última palabra)*, es ciertamente poco convencional, como habría de ser cualquier argumento acerca de la propia argumentación.[5] Nagel lo comparaba al argumento de Descartes en virtud del cual nuestra propia existencia es la única cosa de la que no podemos dudar, porque el hecho mismo de preguntarnos si existimos presupone la existencia de un preguntador. El hecho mismo de cuestionar el concepto de *razón* utilizando la razón presupone la validez de la razón. Debido a esta inconvencionalidad, no es muy correcto decir que deberíamos «creer en» la razón o «tener fe en» la razón. Como señala Nagel, eso es «un pensamiento de más» *(one thought too many)*. Los masones estaban en lo cierto: deberíamos *seguir* la razón.

Ahora bien, los argumentos en defensa de la verdad, la objetividad y la razón pueden resultar irritantes, pues parecen peligrosamente arrogantes: «¿Quién demonios eres *tú* para afirmar que posees la verdad absoluta?». Pero la defensa de la racionalidad no consiste en eso. El psicólogo David Myers ha dicho que la esencia de la creencia monoteísta es: (1) hay un Dios y (2) no soy yo (ni tampoco eres tú).[6] El equivalente secular es: (1) hay una verdad objetiva y (2) yo no la conozco (ni tú tampoco). La misma humildad epistémica es aplicable a la racionalidad que conduce a la verdad. La racionalidad perfecta y la verdad objetiva son aspiraciones que ningún mortal puede afirmar jamás haber alcanzado. Pero la convicción de que estas existen ahí afuera nos autoriza a desarrollar reglas que todos podamos cumplir y que nos permitan aproximarnos colectivamente a la verdad de maneras que resultan imposibles para cualquiera de nosotros individualmente.

Las reglas se diseñan para dejar de lado los sesgos que se interponen en el camino de la racionalidad: las ilusiones cognitivas incorporadas en la naturaleza humana, así como el fanatismo, los prejuicios, las fobias y los -*ismos* que infectan a los miembros de una raza, una clase, un género, una sexualidad o una civilización. Estas reglas incluyen los principios del

pensamiento crítico y los sistemas normativos de la lógica, de la probabilidad y del razonamiento empírico que se explicarán en los próximos capítulos. Son implementadas entre personas de carne y hueso por instituciones sociales que evitan que los individuos impongan sus egos, sesgos o engaños a todos los demás. «Debe hacerse que la ambición contrarreste la ambición», escribió James Madison acerca de los controles y contrapesos en un Gobierno democrático, y así es como otras instituciones conducen hacia la verdad desinteresada a las comunidades de individuos lastrados por los sesgos y la ambición. A título de ejemplo, cabe mencionar el sistema contradictorio en derecho, la revisión por pares en ciencia, la edición y verificación de datos en periodismo, la libertad de cátedra en las universidades y la libertad de expresión en la esfera pública. La discrepancia es necesaria en las deliberaciones entre mortales. Como dice el dicho, cuanto más discrepemos, más probabilidades habrá de que al menos uno de nosotros tenga razón.

AUNQUE NUNCA PODAMOS *DEMOSTRAR* que el razonamiento sea sensato o que la verdad pueda ser conocida (ya que para ello necesitaríamos asumir la sensatez de la razón), podemos avivar nuestra confianza en que lo son. Cuando aplicamos la razón a la razón misma, descubrimos que no se trata tan solo de un impulso visceral inarticulado, de un misterioso oráculo que susurra verdades a nuestro oído. Podemos exponer las reglas de la razón, y destilarlas y depurarlas en modelos normativos de lógica y probabilidad. Incluso podemos implementarlas en máquinas que reproducen y exceden nuestras facultades racionales. Los ordenadores son literalmente lógica mecanizada y sus circuitos más pequeños se denominan *puertas lógicas*.

Otra garantía de que la razón es válida es que *funciona*. La vida no es un sueño en el que aparecemos en lugares inconexos y ocurren cosas inauditas sin orden ni concierto. Al escalar el muro, Romeo consigue tocar realmente los labios de Julieta. Y, al desplegar la razón en otros sentidos, llegamos a la Luna, inventamos los teléfonos inteligentes y extinguimos la viruela. La disposición del mundo a cooperar cuando aplicamos la razón es un fuerte indicio de que la racionalidad llega realmente a verdades objetivas.

Y, en última instancia, incluso los relativistas que niegan la posibilidad de la verdad objetiva e insisten en que todas las afirmaciones son meramente las narrativas de una cultura carecen del coraje de sus convicciones. Los antropólogos culturales o los eruditos literarios que declaran que las verdades de la ciencia son meramente los relatos de una cultura seguirán tratando la infección de su hijo con antibióticos prescritos por un médico en lugar de con un canto curativo entonado por un chamán. Y aunque el relativismo se adorna a menudo con un halo moral, las convicciones morales de los relativistas dependen de un compromiso con la verdad objetiva. ¿Fue la esclavitud un mito? ¿Fue el Holocausto solamente una de las muchas narrativas posibles? ¿Es el cambio climático una construcción social? ¿O son el sufrimiento y el peligro que definen estos acontecimientos verdaderamente reales: afirmaciones que sabemos que son ciertas en virtud de la lógica, las evidencias y la erudición objetiva? Ahora los relativistas dejan de ser tan relativos.

Por el mismo motivo, no caben las concesiones en lo que atañe a la racionalidad y la justicia social o a cualquier otra causa moral o política. La búsqueda de justicia social parte de la creencia de que ciertos grupos están oprimidos y otros son privilegiados. Estas son afirmaciones fácticas y pueden estar equivocadas (como insisten los propios defensores de la justicia social en respuesta a la afirmación de que son los hombres blancos heterosexuales quienes están oprimidos). Afirmamos estas creencias porque la razón y las evidencias sugieren que son verdaderas. Y, a su vez, la búsqueda está guiada por la creencia de que son necesarias ciertas medidas para rectificar esas injusticias. ¿Es suficiente con igualar las condiciones? ¿O acaso las injusticias pasadas han dejado a ciertos grupos en una situación de desventaja, que solo puede enmendarse mediante políticas compensatorias? ¿Servirían medidas meramente particulares para tranquilizar conciencias sin mejorar la situación de los grupos oprimidos? ¿Empeorarían las cosas? Los defensores de la justicia social necesitan conocer las respuestas a estas preguntas, y la razón es la única manera de que podamos saber algo acerca de cualquier cosa.

Cierto es que la peculiar naturaleza de los argumentos en pro de la razón siempre deja una laguna. Al introducir la defensa de la razón he escrito: «Siempre y cuando las personas estén argumentando y persuadiendo...», pero ese «siempre y cuando» no es baladí. Quienes denigran la racionalidad pueden negarse a participar en el juego. Pueden decir: «No tengo que admitir ante ti que mis creencias están justificadas. Tus demandas de argumentos y pruebas muestran que tú eres parte del problema». En lugar de sentir necesidad alguna de persuadir, las personas que están seguras de tener razón pueden imponer sus creencias por la fuerza. En las teocracias y las autocracias, las autoridades censuran, encarcelan, exilian o queman a aquellos que tienen opiniones equivocadas. En las democracias la fuerza es menos brutal, pero la gente sigue hallando medios para imponer una creencia en lugar defenderla con argumentos. Las universidades modernas, curiosamente (puesto que su misión consiste en evaluar ideas), han estado a la vanguardia de la búsqueda de formas de reprimir opiniones, entre las que se incluyen retirar la invitación y acallar a los oradores, apartar de las aulas a los profesores controvertidos, revocar ofertas de empleo y apoyo, expurgar artículos polémicos de los archivos y catalogar las diferencias de opinión como hostigamiento y discriminación punibles.[7] Responden como Ring Lardner recordaba que hacía su padre cuando el escritor era un niño: «Cállate», explicaba.

Entonces, si sabes que estás en lo cierto, ¿por qué *deberías* tratar de persuadir a los otros mediante la razón? ¿Por qué no limitarte a fortalecer la solidaridad en el seno de tu coalición y movilizarla para luchar por la justicia? Una de las razones es que estarías invitando a hacer preguntas tales como las siguientes: «¿Eres infalible? ¿Estás *seguro* de que estás en lo cierto acerca de *todo*? En tal caso, ¿qué es lo que te diferencia de tus oponentes, que también están seguros de tener razón? ¿Y de las autoridades a lo largo de la historia, que insistían en que estaban en lo cierto, pero que hoy sabemos que se equivocaban?». Si tienes que silenciar a aquellos que discrepan de ti, ¿quiere eso decir que careces de buenos argumentos para demostrar por qué están equivocados? La incriminatoria falta de respuestas

convincentes a semejantes preguntas podría enajenar a aquellos que no han tomado partido, incluidas las generaciones cuyas creencias no están grabadas en piedra.

Y otra razón para no eludir la persuasión es que no habrás dejado a aquellos que discrepan de ti otra alternativa que participar en tu juego y contraatacarte con la fuerza en lugar de los argumentos. Puede que estos sean más fuertes que tú, si no ahora, sí en algún punto del futuro. En ese momento, cuando seas el único anulado, será demasiado tarde para afirmar que tus ideas deberían tomarse en serio en virtud de sus méritos.

¿Dejar de tener sentido?

¿Debemos seguir *siempre* a la razón? ¿Necesito un argumento racional para enamorarme, para querer a mis hijos o para disfrutar de los placeres de la vida? ¿No está bien algunas veces hacer locuras y tonterías, y dejar de tener sentido? Si la racionalidad es algo tan estupendo, ¿por qué la asociamos con una adusta falta de alegría? ¿Estaba en lo cierto el profesor de Filosofía en la obra teatral *Jumpers*, de Tom Stoppard, en su respuesta a la afirmación de que «la Iglesia es un monumento a la irracionalidad»?

> ¡La Galería Nacional es un monumento a la irracionalidad! ¡Todas las salas de conciertos son un monumento a la irracionalidad! ¡Y también lo es un jardín bien cuidado, o el favor de un amante o un albergue para perros callejeros...! ¡Si la racionalidad fuera el criterio para permitir la existencia de las cosas, el mundo sería un gigantesco campo de soja![8]

El resto de este capítulo acepta el desafío del profesor. Veremos que aunque la belleza, el amor y la bondad no son literalmente racionales, tampoco son exactamente irracionales. Podemos aplicar la razón a nuestras emociones y a nuestra moralidad, y existe incluso una racionalidad de orden superior que nos dice cuándo puede ser racional ser irracional.

El profesor de Stoppard puede haberse dejado engañar por el célebre argumento de David Hume de que «la razón es y solo debe ser esclava de las pasiones, y no puede pretender otro oficio que el de servirlas y obedecerlas».[9] Hume, uno de los filósofos más realistas y prácticos de la historia del pensamiento occidental, no estaba aconsejando a sus lectores

que actuaran sin pensar, que viviesen el momento o que se enamorasen perdidamente del señor equivocado.[10] Estaba defendiendo la idea lógica de que la razón es el medio para un fin y no puede decirnos cuál debería ser el fin, ni siquiera que hemos de perseguirlo. Con *pasiones* se estaba refiriendo a la fuente de esos fines: los gustos, los deseos, los impulsos, las emociones y los sentimientos programados en nosotros, sin los cuales la razón no tendría objetivos para hallar la forma de alcanzarlos. Se trata de la distinción entre pensar y querer, entre creer algo que consideras verdadero y desear algo que quieres conseguir. Su tesis estaba más próxima a «sobre gustos no hay nada escrito» que a «si te sienta bien, hazlo».[11] No es ni racional ni irracional preferir la tarta de chocolate a la de queso. Y no es irracional en modo alguno cuidar un jardín, enamorarse, cuidar de los perros callejeros, ir de fiesta como si fuese 1999 o bailar bajo el cielo de diamante con una mano saludando libre.[12]

Ahora bien, la impresión de que la razón puede oponerse a las emociones ha de venir de algún sitio; sin duda, no se trata solamente de un error lógico. Nos mantenemos a distancia de los exaltados, imploramos a las personas que sean razonables y lamentamos diversas aventuras amorosas, arrebatos y actos irreflexivos. Si Hume estaba en lo cierto, ¿cómo puede ser verdad asimismo lo contrario de lo que él escribió? A saber: que las pasiones deben ser con frecuencia esclavas de la *razón*.

De hecho, no es difícil reconciliar ambas cosas. Uno de nuestros objetivos puede ser incompatible con los otros. Nuestro objetivo en un momento dado puede ser incompatible con nuestros objetivos en otras ocasiones. Y los objetivos de una persona pueden ser incompatibles con los de otras. Con esos conflictos, no valdrá decir que deberíamos servir y obedecer a nuestras pasiones. Algo tiene que ceder, y ahí es donde la racionalidad ha de arbitrar. Llamamos a las dos primeras aplicaciones de la razón *sabiduría* y a la tercera *moralidad*. Examinemos cada una de ellas.

CONFLICTOS ENTRE OBJETIVOS

Las personas no quieren una sola cosa. Quieren comodidad y placer, pero también quieren salud, la prosperidad de sus hijos, la estima de sus compañeros y un relato satisfactorio de cómo han vivido su vida. Dado que esos objetivos pueden ser incompatibles (la tarta de queso engorda, los hijos desatendidos se meten en problemas y la ambición feroz provoca el desprecio), no siempre puedes conseguir lo que deseas. Unos objetivos son más importantes que otros: la satisfacción más profunda, el placer más duradero, el relato más convincente. Usamos la cabeza para priorizar nuestras metas y perseguir unas a expensas de otras.

De hecho, algunos de nuestros aparentes objetivos ni siquiera son en realidad *nuestros*: son los objetivos metafóricos de nuestros genes. El proceso evolutivo selecciona genes que llevan a los organismos a tener el mayor número posible de descendientes supervivientes en las clases de entornos en los que vivían sus antepasados. Lo hacen dándonos motivos como el hambre, el amor, el miedo, la comodidad, el sexo, el poder y el estatus. Los psicólogos evolucionistas llaman a estos motivos *próximos*, lo cual significa que ingresan en nuestra experiencia consciente e intentamos llevarlos a cabo deliberadamente. Estos contrastan con los motivos *últimos* de la supervivencia y la reproducción, que son los objetivos figurados de nuestros genes: lo que estos dirían que quieren si pudieran hablar.[13]

Los conflictos entre los objetivos próximos y últimos se desarrollan en nuestras vidas como conflictos entre diferentes objetivos próximos. El deseo por una pareja sexual es un motivo próximo, cuyo motivo último es concebir un hijo. Lo hemos heredado porque nuestros más lujuriosos ancestros, por término medio, tenían más descendencia. Sin embargo, concebir un hijo puede no figurar entre nuestros objetivos próximos, por lo que puede que despleguemos nuestra razón para frustrar ese objetivo último utilizando la anticoncepción. Tener una pareja romántica leal a la que no traicionemos y mantener el respeto de nuestros pares son otros objetivos próximos que nuestras facultades racionales pueden perseguir aconsejando a nuestras facultades no tan racionales que eviten las relaciones peligrosas. De manera análoga, perseguimos el objetivo próximo de un cuerpo

saludable y esbelto, ignorando otro objetivo próximo, un postre delicioso, surgido a su vez del objetivo último de acumular calorías en un entorno que escatima la energía.

Cuando decimos que alguien está actuando de forma emocional o irracional, con frecuencia aludimos a malas elecciones en estos intercambios. A menudo se antoja genial en el calor del momento estallar de ira cuando alguien nos enoja. Pero nuestra cabeza más fría puede percatarse de que es preferible contener la furia con el fin de lograr cosas que nos harán sentirnos mejor aún a largo plazo, como una buena reputación y una relación de confianza.

CONFLICTOS ENTRE MARCOS TEMPORALES

Habida cuenta de que no todo sucede a la vez, los conflictos entre objetivos implican con frecuencia metas que se alcanzan en momentos diferentes. Y, a su vez, estas parecen con frecuencia conflictos entre diferentes yos, un yo actual y un yo futuro.[14]

El psicólogo Walter Mischel captó el conflicto en una angustiosa elección que ofreció a niños de cuatro años en un célebre experimento de 1972: una nube de golosina ahora o dos nubes dentro de quince minutos.[15] La vida es un desafío interminable de test de golosinas, dilemas que nos obligan a escoger entre una pequeña recompensa más temprana y una gran recompensa más adelante. Ver una película ahora o aprobar un curso más adelante; comprar una fruslería ahora o pagar el alquiler más adelante; disfrutar cinco minutos de felación ahora o una trayectoria intachable en los libros de historia más adelante.

El dilema de la nube de golosina recibe varios nombres, entre los que se incluyen *autocontrol*, *aplazamiento de la gratificación*, *preferencia temporal* y *descuento del futuro*.[16] Figura en cualquier análisis de la racionalidad porque contribuye a explicar la idea equivocada de que el exceso de racionalidad conduce a una vida estrecha e insulsa. Los economistas han estudiado el fundamento normativo del autocontrol (cuándo *deberíamos* darnos un capricho ahora o esperar a más adelante), ya que este es la base de los tipos de interés, que compensan a las personas por

renunciar a su dinero ahora a cambio de dinero más adelante. Ellos nos han recordado que la elección racional consiste con frecuencia en permitirnos algo ahora: todo depende de cuándo y de cuánto. De hecho, esta conclusión forma parte ya de nuestra sabiduría popular, plasmada en aforismos y chistes.

En primer lugar, más vale pájaro en mano que ciento volando. ¿Cómo sabes que el experimentador cumplirá su promesa y te recompensará por tu paciencia con dos nubes de golosina cuando llegue la hora? ¿Cómo sabes que el fondo de pensiones continuará siendo solvente cuando te jubiles y que el dinero que has ahorrado estará disponible cuando lo necesites? No solo la integridad imperfecta de los fideicomisarios puede penalizar el aplazamiento de la gratificación; también puede hacerlo el conocimiento imperfecto de los expertos. «Todo lo que decían que era malo para ti es bueno para ti», bromeamos, y con la mejor ciencia de la nutrición actual sabemos que muchos placeres que ofrecen los huevos, los camarones y los frutos secos fueron sacrificados en las últimas décadas sin ningún motivo justificado.

En segundo lugar, a la larga todos moriremos. Mañana puedes ser alcanzado por un rayo, en cuyo caso todo el placer que has aplazado para la próxima semana, el próximo año o el próximo decenio se habrá echado a perder. Como aconseja la pegatina de los parachoques: «La vida es corta. Cómete primero el postre».

En tercer lugar, solo se es joven una vez. Puede costarte más contratar una hipoteca en tu treintena que ahorrar y pagar la casa cuando seas un octogenario, pero con la hipoteca consigues vivir en ella todos esos años. Y los años no son solo más numerosos, sino también diferentes. Como me dijo en cierta ocasión mi médico después de una prueba de audición: «La gran tragedia de la vida es que, cuando tienes la edad suficiente para permitirte un buen equipo de sonido, ya no puedes apreciar la diferencia». Esta viñeta expresa una idea similar.



«Ya ves, el problema de hacer cosas para prolongar tu vida es que todos los años extras vienen al final, cuando ya eres viejo.»
The New Yorker © Conde Nast

Estos argumentos se combinan en una historia. Un hombre es condenado a la horca por ofender al sultán y ofrece un trato al tribunal: si le conceden un año, enseñará a cantar al caballo del sultán y se ganará su libertad. Cuando regresa al banquillo de los acusados, otro reo le dice: «¿Estás loco? Solo estás posponiendo lo inevitable. Dentro de un año lo pagarás con creces». El hombre responde: «Me figuro que, en un año, pueden cambiar muchas cosas. Puede que muera el sultán y que el nuevo sultán me indulte. Puede que yo muera, en cuyo caso no habré perdido nada. Puede que muera el caballo y me perdonen la vida. Y, ¿quién sabe? ¡Puede que enseñe a cantar al caballo!».

¿Quiere esto decir que, después de todo, es racional comer la nube de golosina ahora? No del todo; depende de cuánto tengas que esperar y de cuántas nubes consigas por esperar. Dejemos a un lado el envejecimiento y otros cambios, y asumamos en aras de la simplicidad que todos los momentos son iguales. Supongamos que todos los años existe una probabilidad del 1 % de que te alcance un rayo. Eso significa que la probabilidad de que sigas vivo dentro de un año es 0,99. ¿Qué posibilidades

tienes de seguir vivo dentro de dos años? Para que eso suceda, habrás tenido que escapar al rayo un segundo año, con una probabilidad total de 0,99 × 0,99, es decir, $0{,}99^2$ o 0,98 (repasaremos las matemáticas en el capítulo 4). Tres años: 0,99 × 0,99 × 0,99, o $0{,}99^3$ (0,97); diez años: $0{,}99^{10}$ (0,90); veinte años: $0{,}99^{20}$ (0,82) y así sucesivamente, en un descenso exponencial. Por consiguiente, teniendo en cuenta la posibilidad de que nunca llegues a disfrutarla, una nube en mano equivale a nueve décimas partes de nube dentro de una década. Los riesgos adicionales (un experimentador desleal, la posibilidad de que dejen de gustarte las nubes de golosina) alteran las cifras, pero no la lógica. Es racional descontar el futuro *exponencialmente*. Por eso, el experimentador ha de prometer recompensar tu paciencia con más nubes cuanto más tiempo esperes, para pagarte los intereses. Y el interés crece exponencialmente para compensar el declive exponencial del valor actual del futuro para ti.

Esto significa, a su vez, que vivir el presente puede ser irracional de dos maneras. Una es que podemos menospreciar una recompensa futura demasiado abruptamente, poniéndole un precio excesivamente bajo, dada la probabilidad de que sigamos vivos para verlo y el disfrute que nos reportará. La impaciencia puede cuantificarse. Shane Frederick, inventor del test de reflexión cognitiva del capítulo precedente, presentó a sus encuestados unos hipotéticos test de la nube de golosina utilizando recompensas para adultos, y descubrió que una mayoría (especialmente aquellos que se dejaban engañar por las seductoras respuestas incorrectas en los acertijos) prefería tres mil cuatrocientos dólares en ese momento y lugar que tres mil ochocientos dólares un mes después, el equivalente a renunciar a una inversión con un 280 % de rendimiento anual.[17] En la vida real, en torno a la mitad de los estadounidenses próximos a la edad de jubilación no han ahorrado *nada* para su retiro: habían planeado sus vidas como si fuesen a estar muertos para entonces (como les sucedía, de hecho, a la mayor parte de nuestros antepasados).[18] Como le respondió Homer Simpson a Marge cuando esta le advirtió de que se arrepentiría de su conducta: «Ese es un problema para el Homer futuro. ¡Caramba, no envidio a ese tipo!».

El ritmo óptimo al que descontar el futuro es un problema al que no solo nos enfrentamos como individuos, sino también como sociedades, cuando decidimos cuántos fondos públicos deberíamos destinar a beneficiar a nuestros yos mayores y a las generaciones futuras. En todo caso, sí que debemos descontarlo. No es solo que un sacrificio actual sería en vano si un asteroide nos hace terminar como los dinosaurios. Se trata asimismo de que nuestra ignorancia de lo que nos deparará el futuro, incluidos los avances tecnológicos, crece exponencialmente cuanto más remotos son nuestros planes (¿quién sabe?, puede que enseñemos a cantar al caballo). Habría tenido poco sentido que nuestros antepasados se hubieran apretado el cinturón hace un siglo en beneficio nuestro —por ejemplo, desviando dinero de las escuelas y las carreteras para reservar un arsenal de pulmones de acero para prepararse para una epidemia de poliomielitis—, habida cuenta de que somos seis veces más ricos y hemos solucionado algunos de sus problemas, mientras que afrontamos otros nuevos con los que ellos no habrían podido siquiera soñar. Al mismo tiempo, podemos maldecir algunas de sus decisiones cortas de miras, cuyas consecuencias estamos sufriendo en la actualidad, como los entornos esquilmados, las especies extintas y la planificación urbana centrada en los vehículos.

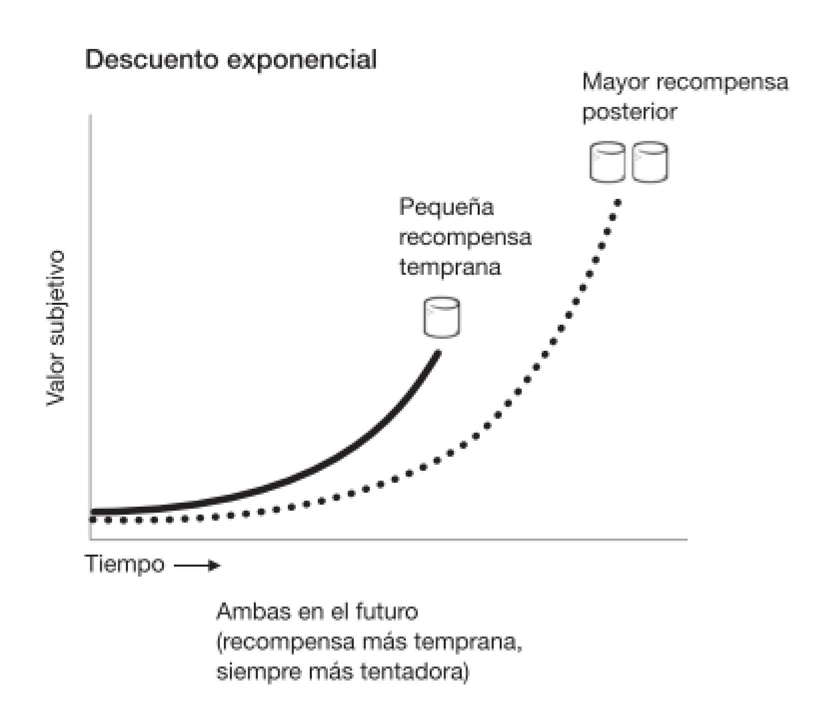
Las decisiones públicas a las que hoy nos enfrentamos, como cuántos impuestos deberíamos pagar sobre el carbono para mitigar el cambio climático, dependen del ritmo al que descontemos el futuro, denominado a veces *tasa de descuento*.[19] Un ritmo del 0,1 %, que refleja solamente la probabilidad de que nos extingamos, significa que valoramos a las generaciones futuras casi tanto como a nosotros mismos, y exige invertir la mejor parte de nuestra renta actual en fomentar el bienestar de nuestros descendientes. Una tasa del 3 %, que asume unos conocimientos y una prosperidad crecientes, demanda posponer la mayor parte del sacrificio a las generaciones que puedan costeárselo con más facilidad. No existe ninguna tasa «correcta», pues depende asimismo de la decisión moral de cómo sopesamos el bienestar de las personas vivas con respecto a las que aún no han nacido.[20] Ahora bien, nuestra conciencia de que los políticos responden a los ciclos electorales más que al largo plazo, y nuestra triste experiencia de no estar preparados para los desastres previsibles, como los huracanes y

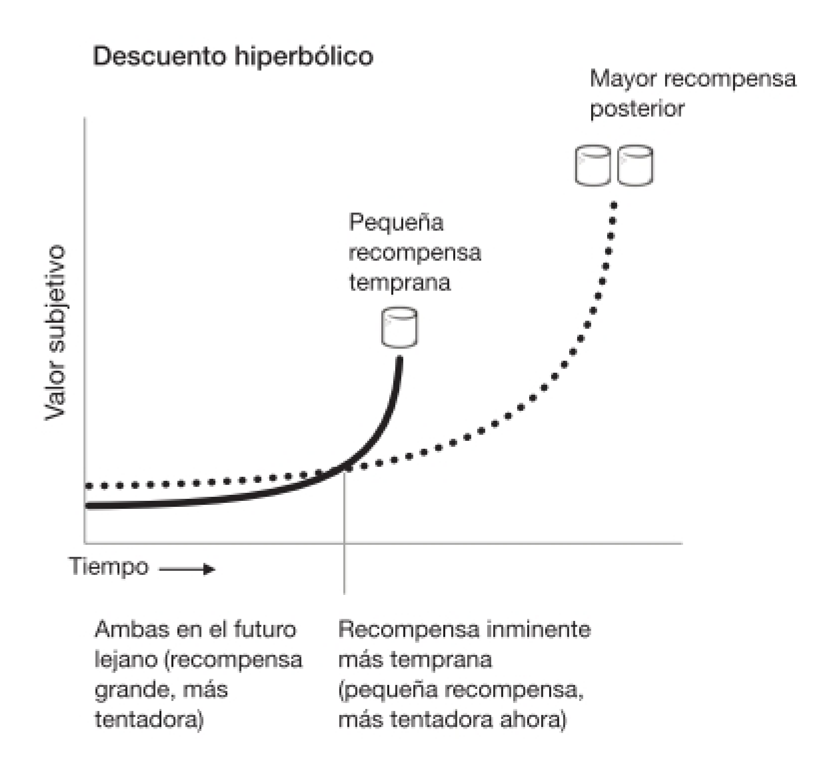
las pandemias, sugieren que nuestra tasa de descuento social es irracionalmente elevada.[21] Dejamos los problemas para el Homer futuro y no envidiamos a ese tipo.

Existe una segunda forma en la que engañamos irracionalmente a nuestros yos futuros, llamada *descuento miope*.[22] Con frecuencia, somos perfectamente capaces de aplazar la gratificación de un yo futuro a otro yo más futuro aún. Cuando el organizador de un congreso envía con antelación un menú para la cena inaugural, es fácil marcar las casillas de las verduras al vapor y la fruta en lugar de la lasaña y la tarta de queso. ¿El pequeño placer de una cena deliciosa dentro de cien días frente al enorme placer de un cuerpo esbelto dentro de ciento un días? ¡No hay punto de comparación! Ahora bien, si el camarero nos tentara con la misma elección en el acto —el pequeño placer de una cena deliciosa en quince minutos frente al enorme placer de un cuerpo esbelto mañana—, invertimos nuestra preferencia y sucumbimos a la lasaña.

La inversión de las preferencias se denomina *miope* o *corta de vista* porque vemos una tentación atractiva que está próxima a nosotros en el tiempo con suma claridad, en tanto que las elecciones remotas son emocionalmente borrosas y (contraviniendo la metáfora oftalmológica) las juzgamos con más objetividad. El proceso racional del descuento exponencial, incluso si el ritmo de descuento es irrazonablemente pronunciado, no puede explicar la inversión, porque si una pequeña recompensa inminente resulta más tentadora que una grande posterior, se antojará más atractiva todavía cuando ambas recompensas se posterguen hacia el futuro (si la lasaña es más apetecible que las verduras al vapor en este momento, la perspectiva de la lasaña sería más atractiva que la perspectiva de las verduras dentro de varios meses). Los científicos sociales dicen que una inversión de las preferencias muestra que el descuento es *hiperbólico*, no en el sentido de ser exagerado, sino de describir una curva llamada hipérbola, que tiene más forma de «L» que de descenso exponencial; comienza con una caída pronunciada y luego se estabiliza. Dos curvas exponenciales a diferentes alturas nunca se cruzan (más tentador ahora, más tentador siempre); dos curvas hiperbólicas pueden cruzarse. Las gráficas de la página siguiente muestran la diferencia

(obsérvese que representan el tiempo absoluto tal como se marca en un reloj o calendario, no el tiempo relativo al ahora, de suerte que el yo que está experimentando las cosas en este momento está recorriendo el eje horizontal y el descuento se muestra en las curvas de derecha a izquierda).





Cierto es que explicar la debilidad de la voluntad a medida que se acerca la recompensa mediante el descuento hiperbólico es como explicar el efecto del hipnótico Zolpidem por su poder dormitivo. Pero la forma de codo de una hipérbola sugiere que esta puede ser en realidad una combinación de dos curvas, una que representa la irresistible atracción de un placer que no puedes sacarte de la cabeza (el aroma de la panadería, una mirada insinuante, el brillo de una sala de exposiciones), y otra que representa una evaluación más fría de costes y beneficios en un futuro hipotético. Los estudios que tientan a los voluntarios en un escáner con versiones adultas de los test de la nube confirman que, al pensar en «golosinas» inminentes y distantes, se activan diferentes patrones cerebrales.[23]

Aunque el descuento hiperbólico no es racional a la manera en que puede serlo el descuento exponencial calibrado (ya que no capta la incertidumbre siempre creciente del futuro), sí que ofrece al yo racional la oportunidad de aventajar al yo impetuoso. La oportunidad puede verse en el segmento más a la izquierda de las hipérbolas, en el momento en que ambas recompensas residen en el futuro remoto, durante el cual la recompensa

grande resulta subjetivamente más atractiva que la pequeña (como debería ser en términos racionales). Nuestro yo más sereno, plenamente consciente de lo que sucederá cuando empiece la cuenta atrás, puede cortar la mitad derecha de la gráfica, no permitiendo que llegue a producirse el cambio a la tentación. El truco se lo explicó Circe a Ulises:[24]

> Primero llegarás a las sirenas, las que hechizan a todos los hombres que se acercan a ellas. Quien acerca su nave sin saberlo y escucha la voz de las sirenas ya nunca se verá rodeado de su esposa y tiernos hijos, llenos de alegría porque ha vuelto a casa; antes bien, lo hechizan estas con su sonoro canto sentadas en un prado donde las rodea un gran montón de huesos humanos putrefactos, cubiertos de piel seca. Haz pasar de largo la nave y, derritiendo cera agradable como la miel, unta los oídos de tus compañeros para que ninguno de ellos las escuche. En cambio, tú, si quieres oírlas, haz que te amarren de pies y manos, firme junto al mástil —que sujeten a este las amarras—, para que escuches complacido la voz de las dos sirenas; y si suplicas a tus compañeros o les ordenas que te desaten, que ellos te sujeten todavía con más cuerdas.

La técnica se denomina *autocontrol de Ulises* y es más efectiva que el extenuante ejercicio de fuerza de voluntad, que es fácilmente derrotado en el momento por la tentación.[25] Durante el precioso interludio antes de que el canto de las sirenas llegue al alcance del oído, nuestras facultades racionales se anticipan a cualquier posibilidad de que nuestros apetitos nos conduzcan a la perdición amarrándonos al mástil, con cuerdas apretadas, suprimiendo la opción de sucumbir. Hacemos la compra cuando estamos saciados y pasamos volando por delante de las patatas fritas y los pasteles, que nos resultan irresistibles cuando estamos hambrientos. Damos instrucciones a nuestro empleador para que diezme nuestro sueldo y reserve una parte para nuestra jubilación, de modo que no haya un excedente a fin de mes para fundírnoslo en unas vacaciones.

De hecho, el autocontrol de Ulises puede subir un nivel y suprimir la opción de tener la opción, o al menos hacerla más difícil de ejercer. Supongamos que la idea de un sueldo íntegro se nos antoja tan tentadora que no somos capaces de rellenar el impreso que autoriza la deducción mensual. Antes de hacer frente a *esa* tentación, podríamos permitir que nuestro empleador tomara la decisión por nosotros (así como otras decisiones que nos beneficien a largo plazo) inscribiéndonos en ahorros obligatorios por defecto: tendríamos que tomar medidas para borrarnos del plan en lugar de para inscribirnos en él. Esta es la base de la filosofía de la

gobernanza caprichosamente llamada *paternalismo libertario* por el experto en derecho Cass Sunstein y el economista comportamental Richard Thaler en su libro *Nudge (Un pequeño empujón [Nudge]).* Estos sostienen que es racional que empoderemos a los Gobiernos y a las empresas para que nos amarren al mástil, aunque con cuerdas flojas en lugar de apretadas. Basándose en las investigaciones sobre el juicio humano, los expertos diseñarían la «arquitectura de decisiones» de nuestros entornos para dificultar que hagamos cosas perjudiciales a la par que tentadoras, como consumir, derrochar y robar. Nuestras instituciones actuarían de manera paternalista, como si supieran lo que es mejor para nosotros, dejándonos la libertad para desatar las cuerdas cuando estemos dispuestos a hacer el esfuerzo (que, de hecho, pocas personas realizan).

El paternalismo libertario, junto con otras «ideas conductuales» extraídas de la ciencia cognitiva, ha llegado a ser cada vez más popular entre los analistas políticos, porque promete resultados más efectivos a bajo costo y sin menoscabo de los principios democráticos. Puede que se trate de la aplicación práctica más importante hasta la fecha de las investigaciones sobre los sesgos cognitivos y las falacias (si bien el enfoque ha sido criticado por otros científicos cognitivos que aducen que los humanos somos más racionales de lo que las investigaciones sugieren).[26]

Ignorancia racional

Mientras Ulises se había hecho amarrar al mástil y había renunciado racionalmente a su opción de *actuar*, sus marineros se taparon los oídos con cera y renunciaron racionalmente a su opción de *saber*. De entrada, esto se antoja desconcertante. Cabría pensar que el conocimiento es poder y que nunca podemos saber demasiado. Al igual que es mejor ser rico que pobre porque, si eres rico, siempre puedes dar tu dinero y ser pobre, podríamos creer que siempre es preferible saber algo, porque siempre puedes decidir no actuar conforme a ello. Sin embargo, en una de las paradojas de la racionalidad, eso resulta no ser cierto. A veces es realmente racional taparse los oídos con cera.[27] La ignorancia puede ser una bendición, y a veces lo que no sabes no puede hacerte daño.

Un ejemplo evidente al respecto es la alerta de *spoiler*. Disfrutamos viendo el desarrollo de una trama, incluido el suspense, el clímax y el desenlace, y podemos optar por no destriparlo conociendo de antemano el final. Los aficionados a los deportes que no puedan ver un partido en tiempo real y planeen ver más tarde una versión grabada se aislarán de todos los medios de comunicación e incluso de otros aficionados que podrían filtrar el resultado de forma sutil. Muchos padres optan por no saber el sexo de su hijo aún no nacido para aumentar el gozo del momento del nacimiento. En tales casos, elegimos racionalmente la ignorancia porque sabemos cómo funcionan nuestras emociones positivas involuntarias y disponemos los acontecimientos para incrementar el placer que nos producen.

Con la misma lógica, podemos entender nuestras emociones negativas y privarnos de la información que prevemos que nos causaría dolor. Muchos consumidores de servicios de genómica saben que sería mejor para ellos permanecer en la ignorancia de si el hombre que dice ser su padre está genéticamente relacionado con ellos. Muchos optan por no saber si han heredado un gen dominante de una enfermedad incurable que mató a uno de sus padres, como el músico Arlo Guthrie, cuyo padre Woody murió de la enfermedad de Huntington. No puede hacerse nada al respecto, y el conocimiento de una muerte temprana y horrible ensombrecería el resto de su vida. En realidad, la mayoría de nosotros nos taparíamos los oídos si un oráculo prometiera revelarnos el día en que moriremos.

Asimismo, evitamos el conocimiento que sesgaría nuestras facultades cognitivas. Los jurados tienen prohibido ver pruebas inadmisibles de habladurías, confesiones forzadas o registros sin orden judicial («la fruta del árbol envenenado»), pues la mente humana es incapaz de ignorarlas. Los buenos científicos piensan lo peor de su propia objetividad y llevan a cabo sus estudios doble ciego, decidiendo no saber qué pacientes tomaron la medicina y cuáles el placebo. Someten sus artículos a la revisión anónima por pares, eliminando así cualquier tentación de tomar represalias después de una evaluación negativa y, en algunas revistas, ocultan sus nombres, con el fin de que los evaluadores no caigan en la tentación de devolver favores o ajustar cuentas.

En estos ejemplos, los agentes racionales optan por la ignorancia a fin de neutralizar sus propios prejuicios nada racionales. Pero a veces decidimos ser ignorantes para impedir que nuestras facultades racionales sean explotadas por adversarios racionales, para asegurarnos de que no puedan hacernos una oferta que no podamos rechazar. Puedes arreglártelas para no estar en casa cuando el mafioso llame a tu puerta con una amenaza o cuando intentan enviarte una citación. El conductor de un camión blindado Brinks está encantado de que se proclame su ignorancia en la pegatina: «El conductor no conoce la combinación de la caja fuerte», porque un atracador no puede amenazarlo de forma creíble para que la revele. Para un rehén es preferible no ver las caras de sus captores, pues ello les deja un incentivo para liberarlo. Incluso los niños que se portan mal saben que es mejor para ellos no encontrarse con la mirada de sus padres.

INCAPACIDAD RACIONAL E IRRACIONALIDAD RACIONAL

La ignorancia racional es un ejemplo de las alucinantes paradojas de la razón explicadas por el politólogo Thomas Schelling en su clásico de 1960 *The Strategy of Conflict (La estrategia del conflicto)*.[28] En determinadas circunstancias puede resultar racional ser no solo ignorante, sino también impotente y, lo más contradictorio de todo, irracional.

En el juego del gallina, que se hizo famoso en el clásico de James Dean *Rebel without a Cause (Rebelde sin causa*), dos conductores adolescentes se aproximan uno a otro a toda velocidad por una carretera estrecha y el primero en virar bruscamente queda humillado (es el «gallina»).[29] Como ambos saben que el otro no quiere morir en un choque frontal, cada uno de ellos puede mantener el rumbo, sabiendo que el otro tiene que virar primero. Por supuesto, cuando ambos son «racionales» en este sentido, es una receta para el desastre (una paradoja de la teoría de juegos sobre la que volveremos en el capítulo 9). ¿Existe entonces una estrategia para ganar en el juego del gallina? Sí: renunciar a tu capacidad de virar bloqueando el volante o colocando un ladrillo en el acelerador y saltar al asiento trasero, no dejando al otro tipo más opción que girar. El jugador

que carece de control gana. Para ser más precisos, el *primer* jugador que carece de control gana: si ambos bloquean sus volantes respectivos simultáneamente...

Aunque el juego del gallina pueda parecer el paradigma de la estupidez adolescente, es un dilema común en el regateo, tanto en el mercado como en la vida cotidiana. Supongamos que estás dispuesto a pagar hasta treinta mil dólares por un coche y sabes que al concesionario le ha costado veinte mil. Cualquier precio entre los veinte mil y los treinta mil dólares resulta ventajoso para ambos, pero, por supuesto, tú quieres que se acerque lo más posible al límite inferior del rango y el vendedor al límite superior. Podrías hacerle una oferta muy baja, sabiendo que prefiere consumar el negocio a desistir, pero él podría pedir un precio muy alto, sabiendo lo mismo. Así pues, el agente comercial conviene en que tu oferta es razonable, pero necesita el visto bueno de su gerente, pero, cuando regresa, dice pesaroso que el gerente es un tipo duro de pelar y ha rechazado el trato. Alternativamente, el cliente conviene en que el precio es razonable, pero necesita el visto bueno de su banco, que se niega a prestarle ese dinero para el coche. El ganador es aquel que tiene las manos atadas. Otro tanto puede suceder en las amistades y en los matrimonios en los que las dos personas prefieren hacer algo juntos a quedarse en casa, pero difieren en lo que más les apetece. La persona supersticiosa, acomplejada o exasperantemente testaruda que descarte categóricamente la elección de la otra se saldrá con la suya.

Las amenazas son otro terreno en el que la falta de control puede proporcionar una paradójica ventaja. El problema de amenazar con atacar, golpear o castigar estriba en que puede resultar costoso cumplir la amenaza y su blanco puede descubrir que se trata de un farol. Para hacerla creíble, quien amenaza debe estar decidido a llevarla a cabo, perdiendo el control que daría a su objetivo la ventaja para amenazarlo a su vez negándose a obedecer. El secuestrador de un avión que lleva un cinturón de explosivos que estalla al mínimo empujón, o los manifestantes que se encadenan a las vías delante de un tren que transporta combustible para una central nuclear no pueden ser ahuyentados de su misión.

El compromiso de cumplir una amenaza puede ser no solo físico, sino también emocional.[30] La pareja sentimental posesiva, exaltada, narcisista o de personalidad límite, o el «hombre de honor» que considera una afrenta intolerable que le falten al respeto y arremete contra los demás independientemente de las consecuencias es alguien con quien más vale no meterse.

La falta de control puede fundirse con la falta de racionalidad. Los terroristas suicidas que creen que serán recompensados en el paraíso no pueden ser disuadidos por la perspectiva de la muerte en la Tierra. Según la teoría del loco en las relaciones internacionales, un dirigente que se considera impetuoso, incluso trastornado, puede coaccionar a un adversario para que haga concesiones.[31] En 1969, Richard Nixon habría ordenado, al parecer, a los bombarderos con armamento nuclear que volasen temerariamente cerca de la Unión Soviética con el fin de asustar a los soviéticos para que presionaran a sus aliados norvietnamitas para que negociasen el final de la guerra de Vietnam. La bravata de Donald Trump en 2017 acerca del uso de su botón nuclear más grande para descargar fuego y furia sobre Corea del Norte podría interpretarse caritativamente como un resurgimiento de la teoría.

Huelga decir que el problema de la estrategia del loco estriba en que ambas partes pueden jugar a ella, diseñando un catastrófico juego del gallina. O bien la parte amenazada puede creer que no tiene otra alternativa que eliminar por la fuerza al loco en lugar de continuar con una negociación infructuosa. En la vida cotidiana, el bando más sensato tiene un incentivo para abandonar una relación con una persona loca y tratar con alguien más razonable. Estas son razones por las que no somos todos locos todo el tiempo (aunque algunos de nosotros nos salgamos con la nuestra en algunas ocasiones).

Las promesas, al igual que las amenazas, tienen un problema de credibilidad que puede requerir una renuncia al control y al propio interés racional. ¿Cómo puede un contratista convencer a un cliente de que pagará cualquier daño, o un prestatario convencer a un prestamista de que devolverá un préstamo, cuando tienen todos los incentivos para faltar a su palabra cuando llegue la ocasión? La solución pasa por depositar una fianza

que perdería o firmar un pagaré que autorice al acreedor a embargar la casa o el coche. Al renunciar a sus opciones, se convierten en socios fiables. En nuestra vida personal, ¿cómo convencemos a un objeto de deseo de que renunciaremos a todos los demás hasta que la muerte nos separe, cuando alguien más deseable todavía puede presentarse en cualquier momento? Podemos anunciar que somos incapaces de elegir racionalmente a alguien mejor porque, para empezar, nunca hemos elegido racionalmente a esa persona: nuestro amor ha sido involuntario, irracional y suscitado por las cualidades únicas, idiosincrásicas e irreemplazables de la persona.[32] «No puedo evitar enamorarme de ti. Estoy loco por ti. Me gusta tu forma de caminar, me gusta tu forma de hablar.»

La paradójica racionalidad de la emoción irracional invita permanentemente a la reflexión y ha inspirado los argumentos de tragedias, wésterns, películas de guerra, de la mafia y de espías, así como los clásicos de la Guerra Fría *Fail Safe (Punto límite)* y *Dr. Strangelove (¿Teléfono rojo? Volamos hacia Moscú)*. Pero en ningún lugar se ha plasmado de manera más sucinta la lógica de la falta de lógica que en la película de cine negro de 1941 *The Maltese Falcon (El halcón maltés*), cuando el detective Sam Spade desafía a los secuaces de Kasper Gutman a que le maten, sabiendo que le necesitan para encontrar el halcón con incrustaciones de joyas. Gutman responde:

> Esa es una actitud, señor, que requiere del más ponderado juicio de ambas partes, porque como usted sabe, señor, en el calor de la acción los hombres olvidan con frecuencia dónde están sus mejores intereses, y se dejan llevar por sus emociones.[33]

Tabú

¿Pueden ciertas ideas ser no solo estratégicamente comprometedoras, sino también malos pensamientos? Este es el fenómeno denominado *tabú*, término derivado de una palabra polinesia que significa «prohibido». El psicólogo Philip Tetlock ha demostrado que los tabúes no son solo costumbres de los isleños del mar del Sur, sino que se hallan activos en todos nosotros.[34]

Una clase de tabú, la *tasa base prohibida*, surge del hecho de que no existen dos grupos de personas (hombres y mujeres, negros y blancos, protestantes y católicos, hindúes y musulmanes, judíos y gentiles) que tengan promedios idénticos en cualquier característica que nos preocupemos de medir. Técnicamente, esas tasas base podrían estar conectadas a fórmulas actuariales, y guiar las predicciones y las políticas pertenecientes a esos grupos. Decir que esa caracterización es arriesgada sería un eufemismo. Examinaremos la moralidad de las tasas base prohibidas en la discusión del razonamiento bayesiano, en el capítulo 5.

Un segundo tipo es el *intercambio tabú*. En la vida, los recursos son finitos y los intercambios inevitables. Dado que no todos valoran todas las cosas por igual, podemos incrementar el bienestar de todo el mundo animando a los individuos a intercambiar algo que sea menos valioso para ellos por algo que les parezca más valioso. Pero este hecho económico se ve contrarrestado por otro psicológico: las personas tratan ciertos recursos como sacrosantos y les ofende la posibilidad de que estos sean intercambiados por productos vulgares, como dinero o comodidad, aun cuando todos ganen.

Un ejemplo de ello es la donación de órganos.[35] Nadie necesita sus dos riñones, mientras que cien mil estadounidenses necesitan uno con desesperación. Esa necesidad no es cubierta ni por los donantes póstumos (incluso cuando el Estado los empuje a acceder, convirtiendo la donación en la opción por defecto) ni por los altruistas vivos. Si a los donantes sanos se les permitiese vender sus riñones (y el Gobierno concediese vales a las personas que no pudieran permitirse pagar), a muchos se les evitaría el estrés financiero, a otros muchos se les evitaría la discapacidad y la muerte, y nadie saldría perjudicado. Sin embargo, la mayoría de las personas no solo se oponen a ese plan, sino que la idea misma les resulta ofensiva. En vez de ofrecer argumentos en contra de ella, se sienten insultados cuando les plantean la posibilidad. Cambiar la recompensa del vil metal a los vales saludables (para educación, asistencia sanitaria o jubilación, pongamos por caso) suaviza la ofensa, pero no la elimina. Análoga es la indignación

cuando se pregunta si deberían existir mercados subvencionados para actuar como jurado, el servicio militar o la adopción de niños dados, ideas ocasionalmente rumoreadas por los traviesos economistas libertarios.[36]

Nos enfrentamos a los intercambios tabúes no solo en políticas hipotéticas, sino en decisiones presupuestarias cotidianas. Un dólar invertido en salud o en seguridad —un paso elevado para peatones, la limpieza de residuos tóxicos— es un dólar no invertido en educación, parques, museos o pensiones. No obstante, los editorialistas no sienten bochorno alguno en hacer absurdas declaraciones como «ninguna cantidad de dinero es excesiva para invertir en *x*» o «no podemos poner un precio a *y*» en lo concerniente a bienes sagrados como el medioambiente, los niños, la sanidad o las artes, como si estuvieran dispuestos a cerrar escuelas para financiar depuradoras de aguas residuales o viceversa. Poner un precio en dólares a una vida humana es repugnante, pero también es inevitable, porque de lo contrario los responsables políticos pueden despilfarrar en causas sentimentales o en proyectos para ganar votos en detrimento de peligros más acuciantes. En Estados Unidos, el precio actual de una vida humana oscila entre los siete y los diez millones de dólares cuando se trata de pagar por la seguridad (aunque los planificadores aceptan de buen grado que el precio se entierre en densos documentos técnicos). Cuando se trata de pagar por la salud, el precio es sumamente variable, una de las razones por las que el sistema estadounidense de salud es tan caro e ineficiente.

Con el fin de mostrar que el mero hecho de *pensar* en los intercambios tabúes se percibe como moralmente corrosivo, Tetlock presentaba a los participantes en su experimento el escenario del administrador de un hospital que se enfrenta a la disyuntiva de gastar un millón de dólares para salvar la vida de un niño enfermo o ahorrarlo para gastos generales del hospital. La gente condenaba al administrador si este daba muchas vueltas al asunto en lugar de tomar una decisión instantánea. Hacía el juicio opuesto, valorando el pensamiento por encima de los reflejos, cuando el administrador lidiaba con un intercambio trágico en lugar de uno tabú: dedicar el dinero a salvar la vida de un niño o de otro.

El arte de la retórica política consiste en ocultar, disfrazar con eufemismos o reformular los intercambios tabúes. Los ministros de finanzas pueden llamar la atención sobre las vidas que salvará una decisión presupuestaria e ignorar las vidas que cuesta. Los reformadores pueden redefinir una transacción desplazando a un segundo plano el toma y daca: los defensores de las mujeres de los barrios chinos hablan de trabajadoras sexuales que ejercen su autonomía en lugar de prostitutas que venden su cuerpo; los anunciantes de seguros de vida (antaño un tabú) describen la medida como un sostén que protege a una familia en vez de como un cónyuge que apuesta porque el otro morirá.[37]

La tercera clase de tabú de Tetlock es el *contrafáctico herético*. Inherente a la racionalidad es la capacidad de ponderar qué *sucedería* si alguna circunstancia *no* fuera verdadera. Esto es lo que nos permite pensar en leyes abstractas en lugar de en el presente concreto, para distinguir la causalidad de la correlación (capítulo 9). La razón por la que decimos que el gallo no causa la salida del sol, aun cuando lo uno siempre viene seguido de lo otro, es que *si* el gallo *no* hubiera cantado, el sol habría salido igualmente.

No obstante, a la gente se le antoja con frecuencia inmoral dejar vagar su mente por ciertos mundos de fantasía. Tetlock preguntaba a los individuos: «Y si José hubiera abandonado a María cuando Jesús era un niño, ¿habría llegado a ser este tan carismático y seguro de sí mismo?». Los cristianos devotos se negaban a responder. Algunos musulmanes devotos son más susceptibles todavía. Cuando Salman Rushdie publicó en 1988 *The Satanic Verses (Los versos satánicos*), una novela que relataba la vida de Mahoma en un mundo contrafáctico en el que algunos de los mandamientos de Alá provenían realmente de Satán, el ayatolá Jomeini emitió una fetua llamando a asesinarlo. En caso de que esta mentalidad se te antoje primitiva y fanática, intenta jugar a este juego en tu próxima cena: «Por supuesto, ninguno de nosotros seríamos jamás infieles a nuestras parejas. Pero supongamos, como una pura hipótesis, que lo fuésemos. ¿Quién sería vuestro amante adúltero?». O prueba este otro: «Por descontado, ninguno de nosotros tiene ni un ápice de racista. Pero imaginemos que lo fuéramos:

¿contra qué grupos tendríais prejuicios?». Una pariente mía fue arrastrada a este juego en cierta ocasión y dejó a su novio cuando este respondió: «Los judíos».

¿Cómo podría ser racional condenar los meros pensamientos, una actividad que, por sí misma, no puede afectar al bienestar de los habitantes del mundo? Tetlock observa que no solo juzgamos a las personas por lo que *hacen*, sino también por quiénes *son*. Una persona que es capaz de albergar ciertas hipótesis, incluso si nos ha tratado bien hasta ahora, podría apuñalarnos por la espalda o traicionarnos si alguna vez se presentara la tentación. Imagínate que alguien te preguntase: «¿Por cuánto dinero venderías a tu hijo? ¿O tu amistad, o tu ciudadanía o tus favores sexuales?». La respuesta correcta es negarte a responder; mejor aún, sentirte ofendido por la pregunta. Al igual que sucede con los hándicaps racionales en los tratos, las amenazas y las promesas, un hándicap en la libertad mental puede suponer una ventaja. Confiamos en aquellos que son constitutivamente incapaces de traicionarnos o de traicionar nuestros valores, no en quienes simplemente han decidido no hacerlo hasta ahora.

Moralidad

Otro ámbito excluido a veces de lo racional es la moral. ¿Podemos llegar a deducir lo que es bueno o malo? ¿Podemos confirmarlo con datos? No resulta evidente cómo podríamos hacer tal cosa. Muchos creen que «no podemos derivar un *deber ser* de un *ser*». La conclusión se le atribuye en ocasiones a Hume, con una justificación similar a su argumento de que la razón debe ser esclava de las pasiones. Es célebre su declaración de que «no es contrario a la razón preferir la destrucción del mundo entero a un rasguño en mi dedo».[38] No es que Hume fuese un sociópata insensible. En justa correspondencia, proseguía: «No es contrario a la razón que yo prefiera mi ruina total con tal de evitar el menor sufrimiento a un indio o a cualquier persona totalmente desconocida». Las convicciones morales parecerían depender de preferencias no racionales, al igual que el resto de las pasiones. Esto concordaría con la observación de que lo que se considera moral e inmoral varía de una cultura a otra, como el vegetarianismo, la blasfemia, la

homosexualidad, el sexo prematrimonial, las palizas, el divorcio y la poligamia. Asimismo varía de un periodo histórico a otro dentro de nuestra propia cultura. En los viejos tiempos, un vistazo a las medias se veía como algo escandaloso.[*]

Los enunciados morales han de distinguirse en efecto de los lógicos y de los empíricos. Los filósofos de la primera mitad del siglo XX se tomaron en serio el argumento de Hume y lidiaron con el posible significado de los enunciados morales si no son lógicos ni declaraciones de hechos empíricos. Algunos llegaron a la conclusión de que «*x* es malo» significa poco más que «*x* es contrario a las reglas», «*x* me desagrada» o, incluso, «¡*x*, buu!».[39] Stoppard se burla de esto en *Jumpers*, cuando un inspector que está investigando un tiroteo es informado por el protagonista acerca de la tesis de un colega filósofo, en virtud de la cual los actos inmorales son «no *pecaminosos*, sino simplemente antisociales». El atónito inspector le pregunta: «¿Piensa acaso que no hay nada *malo* en matar a la gente?». George responde: «Bueno, dicho así, por supuesto... Pero, *filosóficamente*, él piensa que en realidad no es intrínsecamente malo en sí mismo, no».[40]

Al igual que el incrédulo inspector, muchas personas no están dispuestas a reducir la moralidad a convención o a gustos. Cuando decimos «el Holocausto es malo», ¿acaso nuestras facultades racionales no nos dejan forma alguna de diferenciar esa convicción de «no me gusta el Holocausto» o «mi cultura desaprueba el Holocausto»? ¿No es la posesión de esclavos ni más ni menos racional que llevar un turbante, una kipá o un velo? Si una niña está mortalmente enferma y sabemos que existe un medicamento que podría salvarla, ¿no es más racional administrarle el medicamento que negárselo?

Al enfrentarse a esta intolerable implicación, algunas personas confían en conferir un poder superior a la moralidad. Para eso sirve la religión, dicen incluso muchos científicos, como Stephen Jay Gould.[41] Pero Platón se despachó con rapidez este argumento dos mil cuatrocientos años atrás en el *Eutifrón*.[42] ¿Es algo moral porque Dios lo ordena u ordena Dios algunas cosas porque son morales? Si es cierto lo primero, y Dios no tenía razón alguna para sus mandamientos, ¿por qué habríamos de tomarnos en serio sus caprichos? Si Dios te ordenase torturar y matar a un niño, ¿lo

convertiría eso en bueno? «¡Él jamás haría tal cosa!», podrías objetar. Pero eso nos enfrenta a la otra parte de la disyuntiva. Si Dios tiene buenos motivos para sus mandamientos, ¿por qué no apelamos directamente a esas razones y evitamos al intermediario? (Da la casualidad de que el Dios del Antiguo Testamento sí que ordenaba a la gente que sacrificase a niños con bastante frecuencia.)[43]

De hecho, no es difícil fundamentar la moralidad en la razón. Puede que Hume estuviese técnicamente en lo cierto cuando escribió que no es contrario a la razón preferir un genocidio global a un rasguño en nuestro dedo meñique. Pero sus fundamentos eran tremendamente limitados. Como él observaba, *tampoco* es contrario a la razón preferir que nos sucedan a nosotros mismos cosas malas en lugar de cosas buenas; pongamos por caso, anteponer el dolor, la enfermedad, la pobreza y la soledad al placer, la salud, la prosperidad y la buena compañía.[44] De acuerdo. Pero digamos ahora simplemente (de manera irracional, caprichosa, tozuda, sin ningún buen motivo) que preferimos que nos ocurran cosas buenas a cosas malas. Hagamos una segunda suposición salvaje y loca: que somos animales sociales que vivimos con otras personas, en lugar de Robinson Crusoe en una isla desierta, de suerte que nuestro bienestar depende de lo que otros hagan, como ayudarnos cuando tengamos necesidad y no dañarnos sin un buen motivo.

Esto lo cambia todo. Tan pronto como comenzamos a insistir a otro: «No debes hacerme daño ni dejar que muera de hambre ni dejar que se ahoguen mis hijos», no podemos mantener asimismo «pero yo puedo hacerte daño y dejarte morir de hambre y dejar que tus hijos se ahoguen», y confiar en que este nos tome en serio. Esto obedece al hecho de que, en cuanto te involucro en una discusión racional, ya no puedo insistir en que solo cuentan mis intereses porque yo soy yo y tú no, en mayor medida en que puedo insistir en que el sitio que yo ocupo es un lugar especial del universo porque resulta que yo estoy en él. Los pronombres *yo*, *mí* y *mío* no poseen ningún peso lógico; se invierten con cada turno de una conversación. Y, por consiguiente, cualquier argumento que privilegie mi bienestar sobre el tuyo o el suyo, en igualdad de condiciones, es irracional.

Cuando combinamos el interés propio y la sociabilidad con la *imparcialidad* —la intercambiabilidad de perspectivas—, obtenemos el núcleo de la moralidad.[45] Obtenemos la *regla de oro*, o las variantes que toman nota del consejo de George Bernard Shaw: «No hagas a los demás lo que deseas para ti; pueden tener gustos diferentes». Esto se traduce en la versión del rabino Hillel: «Lo que es odioso para ti, no lo hagas a tu prójimo» (esa es toda la Torá, dijo cuando fue retado a explicarla apoyándose en una sola pierna; el resto es comentario). Versiones de estas reglas se han descubierto de manera independiente en el judaísmo, el cristianismo, el hinduismo, el zoroastrismo, el budismo, el confucianismo, el islamismo, el bahaísmo y otras religiones y códigos morales.[46] Entre ellas se incluye la observación de Spinoza: «Aquellos que están gobernados por la razón no desean para sí mismos lo que tampoco desean para el resto de la humanidad». Y el imperativo categórico de Kant: «Actúa de tal manera que la máxima de tu conducta pueda valer siempre como ley universal». Y la teoría de la justicia de John Rawls: «Los principios de la justicia se escogen tras un velo de ignorancia» (acerca de las particularidades de nuestra vida). De hecho, el principio puede verse en el más fundamental de todos los enunciados de la moralidad, el que utilizamos para enseñar el concepto a los niños pequeños: «¿Qué te parecería que *él* te lo hiciera a *ti*?».

Ninguno de estos enunciados depende del gusto, la costumbre o la religión. Y aunque el interés propio y la sociabilidad no son racionales en sentido estricto, apenas son independientes de la racionalidad. Para empezar, ¿cómo llegan a existir los agentes racionales? A menos que estemos hablando de ángeles racionales incorpóreos, son productos de la evolución con cuerpos y cerebros frágiles y ávidos de energía. Para haber permanecido vivos el tiempo suficiente para participar en una discusión racional, deben haber evitado las heridas y el hambre, y haberse guiado por el placer y el dolor. Además, la evolución actúa sobre las poblaciones, no sobre los individuos, por lo que un animal racional ha de formar parte de una comunidad, con todos los vínculos sociales que lo impelen a cooperar, protegerse y aparearse. En la vida real, los razonadores deben ser corpóreos

y comunitarios, lo cual significa que el interés propio y la sociabilidad forman parte del paquete de la racionalidad. Y de la mano del interés propio y la sociabilidad viene la implicación que llamamos moralidad.

La imparcialidad, el principal ingrediente de la moralidad, no es solamente una sutileza lógica, una cuestión de intercambiabilidad de los pronombres. En la práctica, también hace que, por término medio, todos salgan beneficiados. La vida ofrece muchas oportunidades de ayudar a alguien o de abstenerse de hacerle daño, con un bajo coste para uno mismo (capítulo 8). Por consiguiente, si todo el mundo se compromete a ayudar y a no hacer daño, todos salen ganando.[47] Por supuesto, eso no significa que las personas sean de hecho perfectamente morales, sino tan solo que existe un argumento racional por el que deberían serlo.

La racionalidad acerca de la racionalidad

A pesar de su falta de genialidad, deberíamos seguir a la razón, y lo cierto es que lo hacemos en muchos sentidos no evidentes. El mero hecho de preguntar por qué deberíamos seguir a la razón supone confesar que deberíamos hacerlo. Perseguir nuestras metas y deseos no es lo opuesto a la razón, sino, en última instancia, la razón por la que tenemos la razón. Desplegamos la razón para alcanzar esos objetivos, y también para priorizarlos cuando no pueden lograrse todos a la vez. Entregarse a los deseos en el momento es racional para un ser mortal en un mundo incierto, siempre y cuando los momentos futuros no se descuenten de un modo demasiado abrupto o miope. Cuando eso sucede, nuestro yo racional actual puede ganarle la partida a un yo futuro menos racional restringiendo sus opciones, un ejemplo de la racionalidad paradójica de la ignorancia, la impotencia, la impetuosidad y el tabú. Y la moralidad no permanece al margen de la razón, sino que dimana de ella tan pronto como los miembros de una especie social que actúa en interés propio abordan imparcialmente los deseos que entran en conflicto y se solapan entre ellos.

Toda esta racionalización de lo aparentemente irracional puede suscitar la preocupación de que alguien pudiera retorcer *cualquier* excentricidad o perversidad para que revelase alguna justificación oculta. Pero la impresión

es falsa: a veces lo irracional es simplemente lo irracional. Las personas pueden estar confundidas o engañadas acerca de los hechos. Pueden perder de vista cuáles son los objetivos más importantes para ellas y cómo alcanzarlos. Pueden razonar falazmente o, lo que es más habitual, en búsqueda del objetivo equivocado, como ganar una discusión en lugar de descubrir la verdad. Pueden ponerse en una posición difícil, serrar la rama sobre la que están sentadas, dispararse a sí mismas en el pie, gastar como un marinero borracho, jugar al gallina hasta un desenlace trágico, enterrar la cabeza en la arena, tirar piedras contra su propio tejado y actuar como si estuviesen solas en el mundo.

Al mismo tiempo, la impresión de que la razón siempre tiene la última palabra no es infundada. Pertenece a la naturaleza misma de la razón el poder siempre dar un paso atrás, observar cómo está siendo aplicada o si está siendo incorrectamente utilizada, y razonar acerca de ese éxito o fracaso. El lingüista Noam Chomsky ha argüido que la esencia del lenguaje humano es la *recursividad*: una frase puede contener un ejemplo de sí misma sin límite.[48] Podemos hablar no solo sobre un perro, sino sobre el perro del mejor amigo de la tía de la hermana del hermano de mi madre; podemos comentar no solo que ella sabe algo, sino que él sabe que ella lo sabe, y ella sabe que él sabe que ella lo sabe, *ad infinitum*. La estructura de la frase recursiva no es solo una forma de presumir. No habríamos desarrollado la capacidad de decir frases incrustadas en frases si no tuviéramos la capacidad de pensar ideas incrustadas en ideas.

Y ese es el poder de la razón: esta puede razonar acerca de sí misma. Cuando algo parece disparatado, podemos buscar una explicación para la aparente locura. Cuando un yo futuro podría actuar irracionalmente, un yo presente puede ganarle la partida. Cuando un argumento racional cae en la falacia o la sofistería, un argumento más racional aún lo revela. Y, si discrepas, si piensas que hay un fallo en este argumento, es la razón la que te permite hacerlo.

CAPÍTULO 3

Lógica y pensamiento crítico

El tipo moderno de lector general puede reconocerse en la conversación por la cordialidad con la que asiente a enunciados borrosos e indistintos: si decimos que lo negro es negro, sacudirá la cabeza y le costará creerlo; si decimos que lo negro no es tan negro, responderá: «Exactamente». No vacila [...] en levantarse en una reunión pública y expresar su convicción de que, a veces, y dentro de ciertos límites, los radios de un círculo tienen tendencia a ser iguales; pero, por otra parte, instaría a llevar un poco más lejos el espíritu de la geometría.

GEORGE ELIOT[1]

En el capítulo anterior nos preguntábamos por qué los humanos parecen impulsados por lo que el señor Spock llamaba «estúpidas emociones». En este examinaremos su irritante «falta de lógica». El capítulo trata de la lógica, no en el sentido lato de la racionalidad misma, sino en el sentido técnico de inferir enunciados verdaderos (conclusiones) a partir de otros enunciados verdaderos (premisas). Partiendo de los enunciados «todas la mujeres son mortales» y «Jantipa es una mujer», por ejemplo, podemos deducir que «Jantipa es mortal».

La lógica deductiva es una herramienta potente a pesar del hecho de que solo puede sacar conclusiones que ya están contenidas en las premisas (a diferencia de la lógica inductiva, el tema del capítulo 5, que nos guía en la generalización a partir de las evidencias). Dado que mucha gente está de acuerdo en muchas proposiciones —todas las mujeres son mortales, el cuadrado de ocho es sesenta y cuatro, las piedras caen hacia abajo y no hacia arriba, el asesinato es malo—, todos podemos abrazar el objetivo de llegar a proposiciones nuevas y menos evidentes. Una herramienta con tamaño poder nos permite descubrir nuevas verdades acerca del mundo desde la comodidad de nuestros sillones, así como resolver disputas relativas a las muchas cosas en las que la gente no está de acuerdo. El filósofo Gottfried Wilhelm Leibniz (1646-1716) fantaseaba con que la lógica podía lograr una utopía epistémica:

> La única manera de rectificar nuestros razonamientos es hacerlos tan tangibles como los de los matemáticos, a fin de poder detectar nuestro error de un vistazo y, cuando existan disputas entre las personas, poder decir simplemente: calculemos, sin más preámbulo, para ver quién está en lo cierto.[2]

Puede que hayas advertido que, tres siglos después, seguimos sin resolver las disputas diciendo «calculemos». Este capítulo explicará por qué. Una razón es que la lógica puede ser realmente ardua, incluso para los lógicos, y resulta fácil utilizar incorrectamente las reglas desembocando en *falacias formales*. Otra es que a menudo la gente ni siquiera intenta respetar las reglas y comete *falacias informales*. El objetivo de sacar a la luz estas falacias y convencer a las personas para que renuncien a ellas se denomina

pensamiento crítico. Pero una razón fundamental por la que no nos limitamos a calcular sin más preámbulos es que la lógica, al igual que otros modelos normativos de racionalidad, es una herramienta que resulta apropiada para perseguir determinadas metas con ciertas clases de conocimientos, y es de poca ayuda con otras.

LÓGICA FORMAL Y FALACIAS FORMALES

La lógica se define como *formal* porque no se ocupa de los contenidos de los enunciados, sino de sus *formas*, es decir, de la manera en que están construidos a partir de sujetos, predicados y términos lógicos como *y*, *o*, *no*, *todos*, *algunos*, *si* y *entonces*.[3] Con frecuencia aplicamos la lógica a enunciados cuyo contenido nos resulta relevante, como «el presidente de Estados Unidos será destituido de su cargo al ser acusado y declarado culpable de traición, soborno u otros graves delitos y faltas». Deducimos que, para que un presidente sea destituido, no solo debe ser acusado, sino también declarado culpable, y que no es preciso que sea declarado culpable tanto de traición como de soborno, con una de ambas condenas es suficiente. Pero las leyes de la lógica son de uso general: son aplicables con independencia de que el contenido sea de actualidad, oscuro o incluso disparatado. Fue esta idea, y no la mera extravagancia, la que llevó a Lewis Carroll a crear los silogismos en su manual *Symbolic Logic [Lógica simbólica]* de 1896, muchos de los cuales se siguen utilizando hoy en día en los cursos de lógica. Por ejemplo, a partir de las premisas «un perrito cojo no le diría a usted "gracias" si le ofreciera una comba en préstamo» y «usted le ha ofrecido una comba en préstamo al perrito», podemos deducir que «el perrito no ha dicho "gracias"».[4]

Los sistemas de lógica se formalizan como reglas que permiten deducir nuevos enunciados a partir de otros viejos, sustituyendo unas secuencias de símbolos por otras. El más elemental se denomina cálculo proposicional. El término latino *calculus* significa «piedrecita», y nos recuerda que la lógica consiste en manipular símbolos mecánicamente, sin ponderar su contenido. Las oraciones simples se reducen a variables, como

P y Q, a las que se asigna un valor de verdad, VERDADERO O FALSO. Pueden formarse enunciados complejos a partir de otros simples mediante los conectores lógicos *y*, *o*, *no*, y *si... entonces*.

Ni siquiera necesitamos saber lo que significan en inglés o en castellano los términos conectores. Su significado consiste únicamente en reglas que nos dicen si un enunciado complejo es verdadero dependiendo de que los enunciados simples sean o no verdaderos. Esas reglas se estipulan en las tablas de verdad. La de la izquierda, que define *y*, puede leerse fila a fila como sigue: «Cuando P es VERDADERO y Q es VERDADERO, eso significa que "P y Q" es VERDADERO; cuando P es VERDADERO y Q es FALSO, eso significa que "P y Q" es FALSO», y así sucesivamente para las otras dos filas.

P	Q	P Y Q
VERDADERO	VERDADERO	VERDADERO
VERDADERO	FALSO	FALSO
FALSO	VERDADERO	FALSO
FALSO	FALSO	FALSO

P	Q	P O Q
VERDADERO	VERDADERO	VERDADERO
VERDADERO	FALSO	VERDADERO
FALSO	VERDADERO	VERDADERO
FALSO	FALSO	FALSO

P	NO P
VERDADERO	FALSO
FALSO	VERDADERO

Pongamos un ejemplo. En el bonito encuentro inicial de la tragedia romántica de 1970 *Love Story*, Jennifer Cavilleri explica a su compañero de Harvard Oliver Barrett IV, a quien condescendientemente llamaba Preppie, por qué supone que había estudiado en una escuela preparatoria privada (*prep school*): «Pareces estúpido y rico». Etiquetemos «Oliver es estúpido» como P, y «Oliver es rico» como Q. La primera fila de la tabla de verdad para *y* establece los hechos simples que tienen que ser verdaderos para que el desprecio conjuntivo de Jenny sea verdadero; ha de pensar que Oliver es estúpido y que también es rico. Él protesta (no con total sinceridad): «En realidad soy inteligente y pobre». Supongamos que «inteligente» significa «NO estúpido» y «pobre» significa «NO rico». Entendemos que Oliver la contradice invocando la cuarta fila de la tabla de verdad: si no es estúpido y no es rico, entonces no es «estúpido y rico». Si todo cuanto quería hacer era contradecirla, también podría haber dicho: «En realidad, soy inteligente y

rico» (fila 2) o «En realidad, soy estúpido y pobre» (fila 3). El caso es que Oliver estaba mintiendo; no es pobre, lo cual significa que su afirmación de que es «inteligente y pobre» resulta falsa.

Jenny responde sinceramente: «No, *yo soy* inteligente y pobre». Supongamos que inferimos cínicamente, siguiendo la invitación del guion, que «los estudiantes de Harvard son ricos *o* inteligentes». Esta inferencia no es una deducción, sino una inducción (una generalización falible a partir de la observación), pero dejemos a un lado cómo llegamos a ese enunciado y examinemos el propio enunciado preguntándonos qué lo haría verdadero. Se trata de una disyunción, un enunciado con una *o*, y puede ser verificado conectando nuestro conocimiento acerca de los futuros amantes a la tabla de verdad de la *o* (columna central), siendo P «rico» y Q «inteligente». Jenny es inteligente, aun cuando no sea rica (fila 3), y Oliver es rico, aunque puede ser o no inteligente (filas 1 o 2), por lo que el enunciado disyuntivo acerca de los estudiantes de Harvard, al menos en lo que concierne a estos dos, es verdadero.

La chanza prosigue:

OLIVER: ¿Por qué eres tan inteligente?
JENNY: No iría a tomar café contigo.
OLIVER: Yo no te lo pediría.
JENNY: Por eso eres estúpido.

Completemos la respuesta de Jenny: «Si me pidieras que fuéramos a tomar café, te diría que no». Basándonos en lo que se nos ha dicho, ¿es verdadero ese enunciado? Se trata de un *condicional*, un enunciado formado con un *si* (el antecedente) y un *entonces* (el consecuente). ¿Cuál es su tabla de verdad? Recordemos de la tarea de selección de Wason (capítulo 1) que el único modo de que «si P entonces Q» sea falso es que P sea verdadero mientras que Q es falso («si una carta está etiquetada como correo urgente, ha de tener un sello de diez dólares» significa que no puede haber cartas urgentes sin un sello de diez dólares). He aquí la tabla:

P	**Q**	**SI P ENTONCES Q**
VERDADERO	VERDADERO	VERDADERO
VERDADERO	FALSO	FALSO

FALSO	VERDADERO	VERDADERO
FALSO	FALSO	VERDADERO

Si tomamos la palabra a los estudiantes, Oliver no se lo pediría. En otros términos, P es falso, lo cual significa a su vez que el enunciado *si... entonces* de Jenny es verdadero (filas 3 y 4, tercera columna). La tabla de verdad implica que su confirmación de asistencia es irrelevante: siempre y cuando Oliver nunca se lo pida, ella está diciendo la verdad. Ahora bien, como sugiere el cierre de la coqueta escena, Oliver acaba pidiéndoselo (P pasa de FALSO a VERDADERO) y ella acepta (Q es FALSO). Eso significa que su condicional «si P entonces Q» era falso, como lo son con frecuencia las bromas juguetonas.

La sorpresa lógica con la que nos hemos encontrado aquí (que siempre que el antecedente de un condicional sea falso, el condicional entero es verdadero —siempre que Oliver nunca se lo pida, ella está diciendo la verdad—) pone de manifiesto una de las formas en las que un condicional en lógica difiere de un enunciado con un *si* y un *entonces* en la conversación ordinaria. Generalmente utilizamos un condicional para referirnos a una predicción justificada, basada en una ley causal comprobable, como «si bebes café, permanecerás despierto». No quedamos satisfechos con considerar verdadero el condicional solo porque nunca se haya probado, como «si bebes zumo de nabo, permanecerás despierto», que sería lógicamente verdadero si nunca bebieras zumo de nabo. Queremos que haya motivos para creer que, en las situaciones contrafácticas en las que P es verdadero (bebes en efecto zumo de nabo), NO Q (te quedas dormido) no sucedería. Cuando sabemos que el antecedente de un condicional es falso o necesariamente falso, sentimos la tentación de decir que el condicional es discutible, irrelevante, especulativo o incluso carente de sentido, no que es verdadero. No obstante, en el sentido lógico estipulado en la tabla de verdad, en el que «si P entonces Q» es solo un sinónimo de «NO [P y NO Q]», este es el extraño resultado: «Si los cerdos tuvieran alas, entonces 2 + 2 = 5» es verdadero, y también lo es «si 2 + 2 = 3, entonces 2

+ 2 = 5». Por esta razón, los lógicos usan un término técnico para referirse al condicional en el sentido de la tabla de verdad, designándolo como *condicional material*.

He aquí un ejemplo de la vida real de por qué la diferencia resulta relevante. Supongamos que queremos evaluar a los expertos según la exactitud de sus predicciones. ¿Cómo deberíamos tratar una predicción condicional (de 2008) como «si Sarah Palin llegara a ser presidenta, ¿ilegalizaría todos los abortos»? ¿Consigue reconocimiento el experto porque el enunciado sea verdadero en términos lógicos? ¿O el enunciado en cuestión no debería contar en cualquier caso? En la competición real de pronósticos de la que procede este ejemplo, los evaluadores tenían que decidir qué hacer con tales predicciones, y optaron por no considerarlas predicciones verdaderas: eligieron interpretar el condicional en su sentido cotidiano, no como un condicional material en el sentido lógico.[5]

La diferencia entre *si* en el castellano ordinario y *si* en lógica es solo un ejemplo de cómo los símbolos nemotécnicos que utilizamos como conectores en la lógica formal no son sinónimos de las formas en que se emplean en la conversación, en la que, como todas las palabras, poseen múltiples significados que se desambiguan en el contexto.[6] Cuando escuchamos: «Se sentó y me contó la historia de su vida», interpretamos que la *y* implica que primero hizo lo uno y luego lo otro, aunque lógicamente podría haber sido a la inversa (como en el chiste de otra época: «se casaron y tuvieron un hijo, pero no en ese orden»). Cuando el atracador dice: «La bolsa o la vida», es técnicamente exacto que podrías conservar la bolsa y la vida, porque «P O Q» abarca el caso en que P es verdadero y Q es verdadero. Pero resultaría imprudente insistir en el argumento ante tu asaltante; todo el mundo interpreta la *o* en ese contexto como el conector lógico XOR, «o exclusiva», «P O Q Y NO [P Y Q]». Esa es también la razón por la que, cuando en el menú se ofrece «sopa o ensalada», no discutimos con el camarero que tenemos derecho lógicamente a ambas. Y, en términos técnicos, proposiciones tales como «los chicos serán chicos», «un trato es un trato», «es lo que es» y «a veces un puro es solo un puro» son tautologías vacías, necesariamente verdaderas en virtud de su forma y, por

consiguiente, desprovistas de contenido. Pero las interpretamos como dotadas de sentido; en el último ejemplo (atribuido a Sigmund Freud), que un puro no siempre es un símbolo fálico.

INCLUSO CUANDO LAS PALABRAS se limitan a sus sentidos estrictamente lógicos, la lógica sería un ejercicio menor si consistiera exclusivamente en verificar si los enunciados que contienen términos lógicos son verdaderos o falsos. Su potencia dimana de las reglas de la *inferencia* válida: pequeños algoritmos que nos permiten saltar de premisas verdaderas a una conclusión verdadera. La más famosa se denomina «afirmación del antecedente» o *modus ponens* (las premisas se escriben por encima de la línea y la conclusión por debajo):

SI P ENTONCES Q
P

Q

«Si alguien es una mujer, entonces es mortal. Jantipa es una mujer. Por tanto, Jantipa es mortal.» Otra regla de inferencia válida se denomina «negación del consecuente», «ley de contraposición» o *modus tollens*:

SI P ENTONCES Q
NO Q

NO P

«Si alguien es una mujer, entonces es mortal. La gorgona Esteno es inmortal. Por tanto, la gorgona Esteno no es una mujer.»

Estas son las más célebres, pero en modo alguno las únicas reglas válidas de inferencia. Desde la época en que Aristóteles formalizó por primera vez la lógica hasta finales del siglo XIX, cuando comenzó a matematizarse, la lógica fue básicamente una taxonomía de las diversas maneras en las que podemos o no deducir conclusiones a partir de varios conjuntos de premisas. Por ejemplo, tenemos la adición disyuntiva válida (pero esencialmente inútil):

P

―――――
P O Q

«París está en Francia. Por tanto, París está en Francia o existen los unicornios.» Y tenemos el más útil silogismo disyuntivo o proceso de eliminación:

P O Q
NO P
―――――
Q

«La víctima fue asesinada con un tubo de plomo o con un candelabro. La víctima no fue asesinada con un tubo de plomo. Por tanto, la víctima fue asesinada con un candelabro.» Cuenta una historia que el lógico Sidney Morgenbesser y su novia asistieron a terapia de pareja, durante la cual los dos pendencieros ventilaban interminablemente sus quejas mutuas. El exasperado consejero les dijo finalmente: «Miren, alguien tiene que cambiar». Morgenbesser respondió: «Bueno, yo no voy a cambiar. Y ella no va a cambiar. Así pues, va a tener que cambiar *usted*».

Más interesante todavía es el principio de explosión, también conocido como «a partir de una contradicción, se sigue cualquier cosa».

P
NO P
―――――
Q

Supongamos que crees P: «París está en Francia». Supongamos también que crees NO P: «París no está en Francia». Por adición disyuntiva, podemos pasar de P a «P O Q»: «París está en Francia o existen los unicornios». Entonces, por el silogismo disyuntivo, podemos pasar de «P O Q» y «NO P a Q»: «París no está en Francia. Por tanto, existen los unicornios». ¡Enhorabuena! Acabas de demostrar lógicamente que existen los unicornios. Con frecuencia se atribuyen erróneamente a Ralph Waldo Emerson las palabras: «La consistencia es el duende de las mentes pequeñas». De hecho, escribió sobre una consistencia *estúpida*, que aconsejaba trascender a las «grandes almas», pero, en cualquier caso, el

insulto es dudoso.[7] Si tu sistema de creencias contiene una contradicción, puedes creer cualquier cosa (Morgenbesser dijo en cierta ocasión de un filósofo que no era de su agrado: «Hay un tipo que afirmaba tanto P como NO P, y luego extraía todas las consecuencias»).[8]

La forma en que las reglas de inferencia válidas generan conclusiones absurdas pone de manifiesto un asunto importante acerca de los argumentos lógicos. Un argumento válido aplica correctamente las reglas de inferencia a las premisas. Solo nos dice que, si las premisas son verdaderas, entonces la conclusión ha de ser verdadera. No se compromete con que las premisas sean de hecho verdaderas y, por consiguiente, no dice nada acerca de la verdad de la conclusión. Esto puede contrastar con un argumento sólido, que aplica correctamente las reglas a premisas verdaderas y conduce por tanto a una conclusión verdadera. He aquí un argumento válido: «Si Hillary Clinton gana las elecciones en 2016, entonces Tim Kaine es el vicepresidente en 2017. Hillary Clinton gana las elecciones en 2016. Por tanto, Tim Kaine es el vicepresidente en 2017». No se trata de un argumento sólido, porque de hecho Clinton no ganó las elecciones. «Si Donald Trump gana las elecciones en 2016, entonces Mike Pence es el vicepresidente en 2017. Donald Trump gana las elecciones en 2016. Por tanto, Mike Pence es el vicepresidente en 2017.» Este argumento es tanto válido como sólido.

La presentación de un argumento válido como si fuera sólido es una falacia común. Un político promete: «Si eliminamos el despilfarro y el fraude en la burocracia, podemos bajar los impuestos, aumentar los beneficios y equilibrar el presupuesto. Yo eliminaré el despilfarro y el fraude. Por tanto, vótame y todo irá mejor». Afortunadamente, la gente es capaz de detectar con frecuencia la falta de solidez, y contamos con una familia de réplicas para el sofista que saca conclusiones plausibles a partir de premisas dudosas: «Eso es mucho suponer». «Es pedirle peras al olmo.» «Supongamos una vaca esférica» (entre los científicos, hay un chiste sobre un físico reclutado por un granjero para incrementar la producción de leche). Y luego está mi favorita, el yidis *as di bubbe volt gehat beytsim volt zi gevain mayn zaidah*, «si mi abuela tuviera pelotas, sería mi abuelo».

Por supuesto, muchas inferencias no son siquiera válidas. Los lógicos clásicos recopilaron asimismo un listado de inferencias inválidas o falacias formales, secuencias de enunciados en las que las conclusiones parecen seguirse de las premisas, pero de hecho no es así. La más famosa de estas es la *afirmación del consecuente*: «Si P entonces Q. Q, por tanto P». Si llueve, entonces las calles están mojadas. Las calles están mojadas. Por tanto ha llovido. El argumento no es válido: puede que acabe de pasar un camión de limpieza de calles. Una falacia equivalente es la *negación del antecedente*: «Si P, entonces Q. NO P, por tanto NO Q». No ha llovido, por tanto las calles no están mojadas. Tampoco es válido y por la misma razón. Una manera diferente de expresar esto es que el enunciado «si P entonces Q» no implica su converso, «si Q entonces P», ni tampoco su inverso «si NO P, entonces NO Q».

Pero la gente tiende a afirmar el consecuente, confundiendo «P implica Q» con «Q implica P». Por eso, en la tarea de selección de Wason, tantas de las personas a las que se les pedía que verificasen «si D entonces 3» daban la vuelta a la tarjeta 3. Por eso los políticos conservadores animan a los votantes a pasar de «si alguien es socialista, probablemente sea demócrata» a «si alguien es demócrata, probablemente sea socialista». Por eso los chiflados proclaman que la gente se ha reído de todos los grandes genios de la historia, olvidando que «si eres un genio, entonces se ríen de ti» no implica que «si se ríen de ti, entonces eres un genio». Deberían tener esto presente los gandules que observan que la mayoría de las compañías tecnológicas fueron creadas por desertores de la universidad.

Por fortuna, la gente detecta con frecuencia las falacias. Muchos de los que hemos crecido en los años sesenta todavía nos reímos de los guerreros antidrogas que decían que todos los consumidores de heroína empezaban con la marihuana y, por consiguiente, la marihuana es una puerta de entrada a la heroína. Y luego está Irwin, el hipocondríaco que le dijo a su médico: «Estoy seguro de que tengo una enfermedad hepática». «Eso es imposible —respondió el doctor—. Si tuviera una enfermedad hepática nunca lo sabría; no provoca ningún malestar.» Irwin replica: «¡Esos son exactamente mis síntomas!».

Por cierto, si has prestado mucha atención a la formulación de los ejemplos, habrás advertido que no he preservado sistemáticamente intactas mis P y mis Q, como debería haber hecho si la lógica consiste en manipular símbolos. Antes bien, a veces he alterado sus sujetos, sus tiempos verbales, sus números y sus auxiliares. «Alguien es una mujer» se ha convertido en «Jantipa es una mujer», «si me pidieras» ha alternado con «pida», «la carta ha de tener un sello» ha pasado a «tiene un sello». Estas modificaciones son relevantes: «La carta ha de tener un sello» no contradice literalmente «la carta no tiene un sello». Por eso los lógicos han desarrollado lógicas más potentes que descomponen las P y las Q del cálculo proposicional en partes más pequeñas. Entre estas figuran el cálculo de predicados, que distingue los sujetos de los predicados y *todos* de *algunos*; la lógica modal, que distingue los enunciados que resultan verdaderos en este mundo, como «París es la capital de Francia», de aquellos que son necesariamente verdaderos en todos los mundos, como «2 + 2 = 4»; la lógica temporal, que distingue pasado, presente y futuro; y la lógica deóntica, que se preocupa por el permiso, la obligación y el deber.[9]

RECONSTRUCCIÓN FORMAL

¿Qué utilidad práctica tiene la capacidad de identificar las diversas clases de argumentos válidos e inválidos? Con frecuencia, estos nos permiten desvelar los razonamientos falaces en la vida cotidiana. La argumentación racional consiste en establecer una base común de premisas que todo el mundo acepte como verdaderas, junto con enunciados condicionales que todos convengan en que hacen que una proposición se siga de otra, y luego aplicar reglas válidas de inferencia que generen las implicaciones lógicas, y solamente las lógicas, de las premisas. A menudo, un argumento no alcanza este ideal: emplea una regla de inferencia falaz, como la afirmación del consecuente, o depende de una premisa que nunca se ha enunciado explícitamente, convirtiendo el silogismo en lo que los lógicos denominan un *entimema*. Ahora bien, ningún mortal tiene el tiempo o la capacidad de concentración para exponer hasta la última premisa e implicación de un argumento, por lo que, en la práctica, casi todos los

argumentos son entimemas. Con todo, puede resultar instructivo desentrañar la lógica de un argumento como un conjunto de premisas y condicionales, a fin de detectar mejor las falacias y los supuestos no explicitados. Esto se denomina *reconstrucción formal* y los profesores de Filosofía encargan a veces esta tarea a sus alumnos para que agudicen su razonamiento.

Veamos un ejemplo. Un candidato a las primarias presidenciales demócratas de 2020, Andrew Yang, propugnaba la implementación de una renta básica universal (RBU). He aquí un extracto de su sitio web en el que justifica esta medida; he numerado los enunciados:

> (1) Las personas más inteligentes del mundo predicen que un tercio de los estadounidenses perderán su empleo por causa de la automatización en doce años. (2) Nuestras políticas actuales no están equipadas para gestionar esta crisis. (3) Si los estadounidenses no tienen ninguna fuente de ingresos, el futuro podría ser muy sombrío. (4) Una RBU de mil dólares mensuales —financiada por un impuesto sobre el valor añadido— garantizaría que todos los estadounidenses se beneficiasen de la automatización.[10]

Los enunciados (1) y (2) son premisas fácticas; supongamos que son verdaderas. (3) es un condicional y es incontrovertible. Hay un salto de (3) a (4), pero puede salvarse en dos pasos. Hay un condicional (2a) no formulado (pero razonable): «Si los estadounidenses pierden sus empleos, no tendrán ninguna fuente de ingresos», y está la negación (válida) del consecuente de (3), que conduce a «para que el futuro no sea sombrío, los estadounidenses deben tener una fuente de ingresos». Sin embargo, en un examen detallado descubrimos que el antecedente de (2a), «los estadounidenses perderán sus empleos», nunca se ha enunciado. Todo cuanto tenemos es (1), que las personas más inteligentes *predicen* que estos perderán sus empleos. Para pasar de (1) al antecedente de (2a) necesitamos añadir otro condicional, (1a) «si las personas más inteligentes del mundo predicen algo, esto llegará a ser verdadero». Pero sabemos que este condicional es falso. Einstein, por ejemplo, anunció en 1952 que solo la creación de un Gobierno mundial, P, impediría la inminente autodestrucción de la humanidad, Q («si NO P entonces Q»). No obstante, no se ha creado ningún Gobierno mundial (NO P) y la humanidad no se ha destruido a sí misma (NO Q; al menos asumiendo que *inminente* significa «dentro de

varias décadas»). Inversamente, pueden llegar a ser verdaderas ciertas cosas predichas por personas que no son las más inteligentes del mundo, pero que son expertas en el tema relevante, en este caso la historia de la automatización. Algunos de esos expertos predicen que, por cada empleo perdido a causa de la automatización, se creará otro nuevo que no podemos prever: los operadores de carretillas elevadoras desempleados se reciclarán como técnicos en eliminación de tatuajes, diseñadores de vestuario de videojuegos, moderadores de contenido de medios sociales y psiquiatras de mascotas. En ese caso, el argumento fracasaría; un tercio de los estadounidenses no perderían necesariamente sus empleos, y una RBU sería prematura para prevenir una crisis inexistente.

La intención de este ejercicio no es criticar a Yang, que fue admirablemente explícito en su programa electoral, ni sugerir que confeccionemos un diagrama lógico para cada argumento que consideremos, lo cual resultaría insoportablemente tedioso. Ahora bien, el hábito de la reconstrucción formal, incluso si esta se lleva a cabo parcialmente, puede desvelar con frecuencia inferencias falaces y premisas no formuladas que, de lo contrario, yacerían ocultas en cualquier argumento, por lo que merece la pena cultivarlo.

PENSAMIENTO CRÍTICO Y FALACIAS INFORMALES

Aunque las falacias informales, tales como la negación del antecedente, pueden sacarse a relucir cuando se reconstruye formalmente un argumento, los errores más comunes en el razonamiento no pueden encasillarse de esa manera. En lugar de violar nítidamente la forma de un argumento en el cálculo proposicional, los argumentadores explotan algún señuelo psicológicamente persuasivo, pero intelectualmente espurio. Se trata de las denominadas *falacias informales*, y los entusiastas de la racionalidad les han dado nombres, recopilándolas por docenas y organizándolas (junto con las falacias formales) en páginas web, pósteres y tarjetas, así como en los programas de los cursos de primer año sobre «pensamiento crítico»[11] (no he podido resistirme; véase el índice analítico y de nombres).

Muchas falacias informales dimanan de una característica del razonamiento humano tan arraigada en nosotros que, según los científicos cognitivos Dan Sperber y Hugo Mercier, fue la presión selectiva la que permitió que evolucionase el razonamiento. Nos gusta vencer en las discusiones.[12] En un foro ideal, el ganador de una discusión es aquel que defiende la posición más convincente. Pero pocas personas poseen la paciencia rabínica para reconstruir formalmente una argumentación y evaluar su corrección. La conversación ordinaria mantiene su unidad mediante los vínculos intuitivos que nos permiten conectar los puntos incluso cuando la discusión no alcanza una explicitud talmúdica. Los individuos diestros en el debate pueden explotar estos hábitos para crear la ilusión de que han fundamentado una proposición sobre unos sólidos cimientos lógicos, cuando en realidad esta se halla levitando por los aires.

La más destacada de las falacias informales es el *hombre de paja*, la efigie de un oponente que es más fácil de derribar que la cosa real. «Noam Chomsky afirma que los niños nacen hablando.» «Kahneman y Tversky dicen que los humanos somos imbéciles.» Esta posee una variante en tiempo real practicada por los entrevistadores agresivos: la táctica del *entonces lo que está usted diciendo es*... «Las jerarquías de dominancia son comunes en el mundo animal, incluso en criaturas tan simples como las langostas». «Entonces lo que usted está diciendo es que deberíamos organizar nuestras sociedades a la manera de las langostas».[13]

Al igual que los argumentadores pueden reemplazar a hurtadillas la proposición de un oponente por otra más fácil de atacar, pueden sustituir su propia proposición por otra más fácil de defender. Pueden recurrir al *alegato especial*, por ejemplo, al explicar que la percepción extrasensorial no supera las pruebas experimentales porque es perturbada por las vibraciones negativas de los escépticos. O que las democracias nunca inician las guerras, exceptuando la antigua Grecia, pero esta tenía esclavos, y la Inglaterra georgiana, pero los plebeyos no podían votar, y los Estados Unidos del siglo XIX, pero sus mujeres carecían del derecho al sufragio, y la India y Pakistán, pero estos eran Estados incipientes. Pueden *cambiar las reglas del juego*, exigiendo la «retirada de financiación a la policía», pero explicando a continuación que solo se refieren a la reasignación de parte de

su presupuesto a los servicios de emergencia (los expertos en la racionalidad la designan como falacia de *motte-and-bailey*, en referencia al castillo medieval dotado de una torre estrecha pero inexpugnable en la que uno puede refugiarse cuando los invasores atacan el patio más deseable, pero menos defendible).[14] Pueden afirmar que ningún escocés echa azúcar a su avena y, cuando se enfrentan con Angus, que echa azúcar a su avena, dicen que eso demuestra que Angus no es un verdadero escocés. La falacia del *falso escocés* explica asimismo por qué ningún verdadero cristiano mata jamás, ningún verdadero Estado comunista es represivo y ningún verdadero partidario de Trump aprueba la violencia.

Estas tácticas se confunden prácticamente con la *petición de principio*, falacia informal consistente en suponer aquello que estamos intentando demostrar. Incluye explicaciones circulares, como en el caso de la *virtus dormitiva* de Molière (la explicación de su médico de por qué el opio hace dormir a la gente), y presuposiciones tendenciosas, como en el clásico «¿cuándo dejó usted de golpear a su mujer?». En un chiste, un hombre alardea del melifluo cantor de su sinagoga y otro replica: «¡Ja! Si yo tuviera su voz, sería tan bueno como él».

Siempre cabe mantener una creencia, cualquiera que esta sea, diciendo que la *carga de la prueba* recae sobre aquellos que discrepan. Bertrand Russell respondió a esta falacia cuando fue retado a explicar por qué era ateo en lugar de agnóstico, ya que no podía demostrar que Dios no existe. Respondió: «Nadie puede demostrar que no haya entre la Tierra y Marte una tetera de porcelana describiendo una órbita elíptica».[15] A veces ambas partes incurren en la falacia, conduciendo al estilo de debate llamado *tenis de la carga* («la carga de la prueba recae sobre ti»; «no, la carga de la prueba recae sobre ti»). En realidad, dado que partimos de la ignorancia acerca de todo, la carga de la prueba recae sobre cualquiera que desee demostrar algo (como veremos en el capítulo 5, el razonamiento bayesiano ofrece una manera basada en principios de razonar acerca de quién debería llevar la carga conforme se acumulan los conocimientos).

Otra táctica de distracción se denomina *tu quoque*, locución latina que significa «tú también». Era una de las favoritas de los apologistas de la Unión Soviética en el siglo XX, que presentaban la siguiente defensa de su

represión totalitaria: «¿Qué hay del trato a los negros en Estados Unidos?». En otro chiste, una mujer regresa temprano a casa del trabajo y encuentra a su marido en la cama con su mejor amiga. Sobresaltado, el hombre dice: «¿Qué haces en casa tan temprano?». Ella responde: «¿Qué haces *tú* en la cama con mi mejor amiga?». Él repone con brusquedad: «¡No cambies de tema!».

La apelación de la Yang Gang (el grupo que respaldaba la candidatura demócrata del empresario Andrew Yang) a «las personas más inteligentes del mundo» constituye un ejemplo suave del *argumento de autoridad*. La autoridad deferida es con frecuencia religiosa, como en la canción de góspel y la pegatina de parachoques «Dios lo dijo, yo lo creo, no hay más que hablar». Pero también puede ser política o académica. Las camarillas intelectuales giran a menudo en torno a un gurú cuyos pronunciamientos se convierten en un evangelio popular. Muchas disquisiciones académicas comienzan con un «como nos ha enseñado Derrida...», o Foucault, o Butler, o Marx, o Freud, o Chomsky. Los buenos científicos reniegan de esta forma de hablar, pero a veces ellos son erigidos en autoridades por otros. Con frecuencia recibo cartas que me llaman a capítulo por preocuparme por el cambio climático provocado por los humanos porque, observan, este brillante físico o aquel laureado Nobel lo niegan. Pero Einstein no fue la única autoridad científica cuyas opiniones ajenas a su campo de especialización no eran nada autorizadas. En su artículo «The Nobel disease: When intelligence fails to protect against irrationality» [«La enfermedad de los nobeles: cuando la inteligencia no logra proteger contra la irracionalidad»], Scott Lilienfeld y sus colegas enumeran las extravagantes creencias de una docena de laureados de la ciencia, incluidas la eugenesia, las megavitaminas, la telepatía, la homeopatía, la astrología, el herbalismo, la sincronicidad, la pseudociencia racial, la fusión fría, los tratamientos extravagantes del autismo y la negación de que el sida es causado por el VIH.[16]

Al igual que el argumento de autoridad, la falacia de *subirse al carro* o *efecto arrastre* explota el hecho de que seamos primates sociales y jerárquicos. «La mayoría de las personas que conozco piensan que la astrología es científica, así que algo de cierto debe de haber.» Aunque

puede que no sea verdad que «la mayoría siempre se equivoca», desde luego no siempre tiene razón.[17] Los libros de historia están repletos de manías, castillos en el aire, cazas de brujas y otros extraordinarios delirios populares y locuras de las multitudes.

Otra contaminación de lo intelectual por lo social es el intento de rebatir una idea cargando contra el carácter, los motivos, los talentos, los valores o las ideas políticas de la persona que la defiende. La falacia se denomina argumentación *ad hominem*, contra la persona. Una versión cruda, pero común, es respaldada por Wally en la tira cómica *Dilbert*:



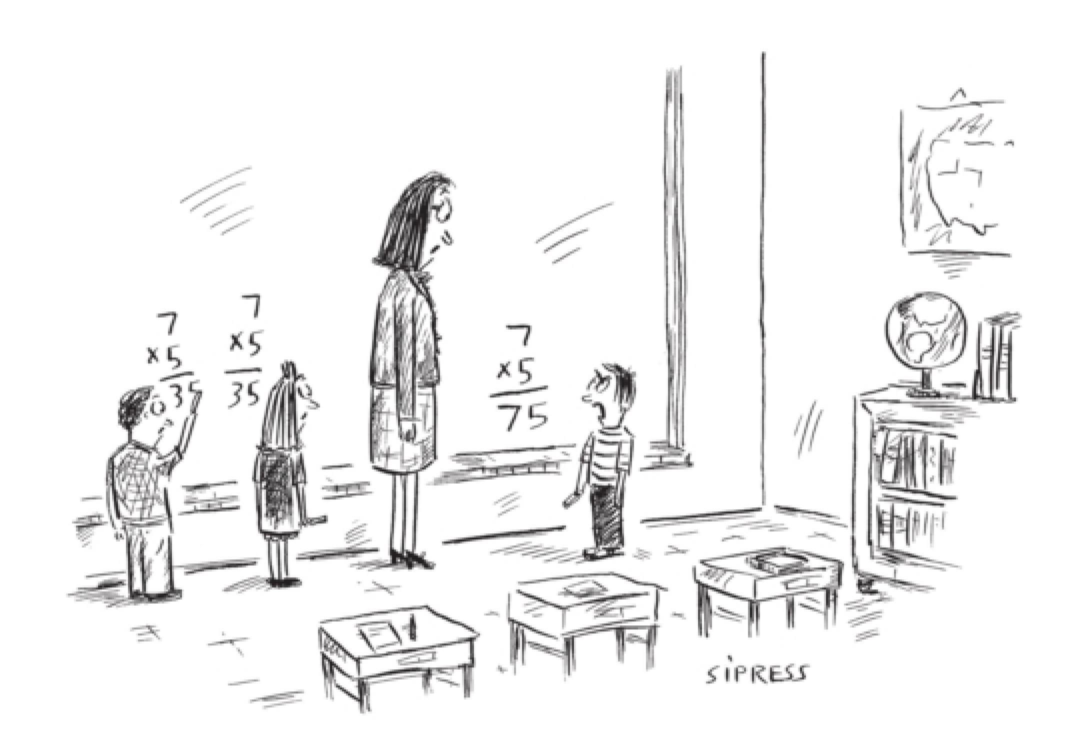
DILBERT © 2020 Scott Adams, Inc. Utilizado con permiso de ANDREWS MCMEEL SYNDICATION.
Todos los derechos reservados.

A menudo la expresión es más gentil, pero no menos falaz. «No debemos tomarnos en serio el argumento de Smith; es un varón blanco heterosexual y enseña en una escuela de negocios.» «El único motivo por el que Jones defiende el cambio climático es que así consigue sus subvenciones, sus becas y sus invitaciones para dar charlas TED.» Una táctica relacionada es la falacia *genética*, que no tiene nada que ver con el ADN, pero está relacionada con las palabras *génesis* y *generar*. Se refiere a la evaluación de una idea no por su verdad, sino por sus orígenes. «Brown consiguió sus datos en el *CIA World Factbook [Libro mundial de hechos de la CIA]*, y la CIA derrocó Gobiernos democráticos en Guatemala e Irán.» «Johnson citó un estudio financiado por una fundación que respaldaba la eugenesia.»

En ocasiones, las falacias *ad hominem* y genéticas se combinan para forjar cadenas de *culpabilidad por asociación*: «La teoría de Williams ha de ser repudiada porque habló en una conferencia organizada por alguien que

publicó un volumen que incluye un capítulo escrito por alguien que dijo algo racista». Aunque nadie puede negar el placer de unirse en contra de un tipo malo, las falacias *ad hominem* y genéticas son genuinamente falaces: las buenas personas pueden profesar malas creencias y viceversa. Por poner un ejemplo punzante, los conocimientos sobre salud pública que permiten salvar vidas, incluida la carcinogenicidad del humo del tabaco, fueron originalmente descubiertos por los científicos nazis, y las compañías tabacaleras estaban encantadas de rechazar el vínculo entre el tabaco y el cáncer porque aquello era «ciencia nazi».[18]

Luego tenemos los argumentos dirigidos directamente al sistema límbico en lugar de a la corteza cerebral. Entre estos figura la *apelación a la emoción*: «¿Cómo puede alguien ver esta foto de los afligidos padres de un niño muerto y decir que las muertes provocadas por la guerra han disminuido?». Y la falacia *afectiva* cada vez más popular, en la que un enunciado puede ser rechazado si resulta «hiriente» o «perjudicial», o si puede provocar «malestar». Aquí vemos a un niño perpetrador de la falacia afectiva:



«Puede que esté mal, pero es lo que yo siento.»
David Sipress/The New Yorker Collection/The Cartoon Bank

Por supuesto, muchos hechos son hirientes: la historia racial de Estados Unidos, el calentamiento global, un diagnóstico de cáncer, Donald Trump. Sin embargo, son hechos a pesar de todo, y hemos de conocerlos lo mejor posible para afrontarlos.

Las falacias *ad hominem*, genéticas y afectivas solían tratarse como errores garrafales para golpearse la frente, o trucos sucios e infames. Los profesores de pensamiento crítico y los entrenadores de debate de secundaria enseñaban a sus alumnos a detectarlas y refutarlas. Sin embargo, en una de las ironías de la vida intelectual moderna, se están convirtiendo en moneda corriente. En grandes sectores del mundo académico y periodístico, las falacias se cultivan con entusiasmo, atacándose o reprimiéndose las ideas porque sus defensores, a veces de hace siglos, tienen manchas y olores desagradables.[19] Esto refleja un cambio en la concepción de la naturaleza de las creencias: de ideas que pueden ser verdaderas o falsas a expresiones de la identidad moral y cultural de una persona. Asimismo denota un cambio en la manera en que los eruditos y los críticos conciben su misión: de buscar el conocimiento a promover la justicia social y otras causas morales y políticas.[20]

Sin lugar a dudas, a veces el contexto de un enunciado es realmente relevante a la hora de evaluar su verdad. Esto puede crear una impresión equivocada de que las falacias informales son aceptables después de todo. Uno puede ser escéptico respecto a un estudio que demuestra la eficacia de un medicamento llevado a cabo por alguien que puede beneficiarse de la medicina, pero advertir un conflicto de intereses no es una falacia *ad hominem*. Uno puede descartar una afirmación basada en la inspiración divina, en la exégesis de los textos antiguos o en la interpretación de las entrañas de una cabra; esta desacreditación no es la falacia genética. Uno puede tomar nota de un cuasi consenso entre los científicos para refutar la afirmación de que debemos ser escépticos en algún asunto porque los expertos discrepan; esto no es la falacia de subirse al carro. Y podemos imponer criterios de prueba más elevados para una hipótesis que exigiría medidas drásticas si fuera cierta; esto no es la falacia afectiva. La diferencia estriba en que, en los argumentos legítimos, uno puede dar *razones* de por qué el contexto de un enunciado debería influir en nuestra creencia en su

verdad o falsedad, o en nuestra manera de obrar al respecto, como la indicación del grado de fiabilidad de la prueba. Con las falacias, uno se rinde a sentimientos que no tienen nada que ver con la verdad de la afirmación.

Así pues, con todas estas falacias formales e informales esperando atraparnos (Wikipedia enumera más de un centenar), ¿por qué no podemos deshacernos de una vez por todas de esta palabrería e implementar el plan de Leibniz para el discurso lógico? ¿Por qué no podemos tornar nuestros razonamientos tan tangibles como los de los matemáticos para ser capaces de detectar nuestros errores de un vistazo? ¿Por qué en el siglo XXI seguimos teniendo discusiones tabernarias, guerras en Twitter, terapias de parejas y debates presidenciales? ¿Por qué no decimos «calculemos» y veamos quién está en lo cierto? No vivimos en la utopía de Leibniz y, al igual que sucede con otras utopías, jamás lo haremos. Existen al menos tres razones para ello.

VERDADES LÓGICAS VERSUS EMPÍRICAS

Uno de los motivos por los que la lógica nunca gobernará el mundo es la distinción fundamental entre proposiciones *lógicas* y *empíricas*, que Hume llamaba *relaciones de ideas* y *cuestiones de hecho*, y los filósofos denominan *analíticas* y *sintéticas*. Para determinar si «todos los solteros son no casados» es verdadero, solo necesitas saber qué significan las palabras (sustituir *soltero* por la frase «varón Y adulto Y NO casado»), y comprobar la tabla de verdad. Ahora bien, para determinar si «todos los cisnes son blancos» es verdadero, tienes que levantarte de tu sillón y observar. Si visitas Nueva Zelanda, descubrirás que la proposición es falsa, porque allí los cisnes son negros.

Se dice con frecuencia que la revolución científica del siglo XVII se inició cuando se empezó a apreciar que los enunciados relativos al mundo físico son empíricos y solamente pueden establecerse mediante la observación, no la argumentación escolástica. Existe una preciosa historia atribuida a Francis Bacon.

> En el año de Nuestro Señor de 1432, surgió una encendida disputa entre los miembros de una comunidad acerca del número de dientes que tiene un caballo. Durante trece días continuó sin descanso la encarnizada polémica. Se sacaron escritos y crónicas, y salió a relucir una erudición tan maravillosa y abrumadora como jamás se había conocido en el lugar. Al comienzo del decimocuarto día, un joven fraile muy piadoso pidió permiso a sus doctos superiores para añadir una idea a la discusión. Y entonces, sin titubeos, y ante el indignado asombro de los presentes, les exhortó a que mirasen en la boca abierta de un caballo, y aclarasen así, de una vez, todas sus dudas. Entonces, al ver su dignidad seriamente herida, se enojaron profundamente y, unidos en un mismo clamor, se abalanzaron sobre él, le golpearon y le arrojaron fuera de la estancia. «Seguramente —dijeron—, Satán ha tentado a este neófito atrevido para que proclame extraños e impíos procedimientos de hallar la verdad, en contra de las enseñanzas de nuestros padres.»

Casi con certeza, este acontecimiento jamás tuvo lugar y resulta dudoso que Bacon lo refiriera.[21] No obstante, la historia capta espléndidamente una de las razones por las que nunca resolveremos nuestras incertidumbres sentándonos a calcular.

Racionalidad formal versus ecológica

Un segundo motivo por el cual el sueño de Leibniz jamás se hará realidad reside en la naturaleza de la lógica formal: al ser *formal*, no ve más que los símbolos y la disposición de estos cuando se despliegan ante el razonador. Es ciega al *contenido* de la proposición: a lo que significan esos símbolos y al contexto y el trasfondo de conocimientos que podrían mezclarse en la deliberación. En sentido estricto, el razonamiento lógico implica el olvido de todo cuanto sabemos. Un estudiante que está haciendo un examen de geometría euclidiana no recibe reconocimiento alguno por sacar unas reglas y medir los dos lados del triángulo con ángulos iguales, por sensato que ello pueda resultar en la vida cotidiana, sino que se le pide que lo demuestre. Análogamente, los estudiantes que hacen los ejercicios de lógica del manual de Carroll no deben distraerse con su conocimiento irrelevante de que los cachorros no pueden hablar. La única razón legítima para concluir que el perrito cojo no diría «gracias» es que eso es lo que está estipulado en el consecuente de un condicional cuyo antecedente es verdadero.

En este sentido, la lógica no es racional. En el mundo en el que hemos evolucionado y en la mayor parte del mundo en el que transcurren nuestros días, carece de sentido ignorar todo lo que sabemos.[22] Sí que tiene sentido

en ciertos mundos no naturales, como los cursos de lógica, los acertijos, la programación informática, los procedimientos jurídicos, la aplicación de la ciencia y la matemática a áreas en las que el sentido común guarda silencio o resulta engañoso. Pero, en el mundo natural, la gente hace muy bien en combinar sus habilidades lógicas con sus conocimientos enciclopédicos, como vimos en el capítulo 1 con los sans. También vimos que, cuando añadimos ciertas clases de verosimilitud a los acertijos, los individuos recurren a sus conocimientos del tema y dejan de avergonzarse. Cierto es que, cuando se les pide que verifiquen «si una tarjeta tiene una D en una cara ha de tener un 3 en la otra», dan la vuelta erróneamente al «3» y olvidan dar la vuelta al «7». Ahora bien, cuando se les pide que se imaginen que son los gorilas de un bar y que verifiquen «si un cliente que está bebiendo alcohol debe ser mayor de veintiún años», saben comprobar las bebidas que tienen delante los adolescentes y pedir la documentación a cualquiera que esté bebiendo cerveza.[23]

El contraste entre la racionalidad *ecológica* que nos permite prosperar en un entorno natural y la racionalidad *lógica* exigida por los sistemas formales es una de las características definitorias de la modernidad.[24] Los estudios de pueblos iletrados por parte de los psicólogos y antropólogos culturales han demostrado que estos están enraizados en la rica textura de la realidad y tienen poca paciencia para los mundos de fantasía con los que están familiarizados los graduados instruidos en Occidente. Veamos una entrevista de Michael Cole a un miembro del pueblo kpelle en Liberia:

> P.: Flumo y Yakpalo siempre beben ron juntos. Flumo está bebiendo ron. ¿Está bebiendo ron Yakpalo?
> R.: Flumo y Yakpalo beben ron juntos, pero cuando Flumo estaba bebiendo el primero, Yakpalo no estaba allí ese día.
> P.: Pero le he dicho que Flumo y Yakpalo siempre beben juntos. Un día Flumo estaba bebiendo ron. ¿Estaba bebiendo ron Yakpalo?
> R.: El día que Flumo estaba bebiendo el ron, Yakpalo no estaba allí.
> P.: ¿Por qué razón?
> R.: Porque ese día Yakpalo se había ido a su granja y Flumo se había quedado en el pueblo.[25]

El kpelle trata la pregunta como una consulta sincera, no como un acertijo lógico. Su respuesta, aunque contaría como un error en un examen, no es en modo alguno irracional: utiliza información relevante para

proponer la respuesta correcta. Los occidentales instruidos han aprendido a jugar al juego de olvidar lo que saben y concentrarse en las premisas de un problema, aun cuando les cueste separar sus conocimientos fácticos de su razonamiento lógico. Muchos insistirán, por ejemplo, en que el siguiente argumento es lógicamente inválido: «Todas las cosas hechas de plantas son saludables. Los cigarrillos están hechos de plantas. Por tanto, los cigarrillos son saludables».[26] Si cambiamos *cigarrillos* por *ensaladas*, confirmarán que es correcto. Los profesores de Filosofía que plantean a sus alumnos experimentos mentales artificiosos, como si resulta permisible arrojar a un hombre gordo sobre un puente para detener un tranvía que circula fuera de control amenazando a cinco operarios en la vía, a menudo se sienten frustrados cuando los estudiantes buscan escapatorias, como gritar a los trabajadores que se quiten de en medio. Sin embargo, esa sería exactamente la manera racional de reaccionar en la vida real.

Las zonas en las que jugamos a juegos formales regidos por reglas (el derecho, la ciencia, los dispositivos digitales, la burocracia) se han expandido en la vida moderna con la invención de potentes fórmulas y reglas ciegas a los contenidos. Pero estas siguen sin estar a la altura de la vida en toda su plenitud. La utopía lógica leibniziana, que requiere una amnesia autoinfligida respecto del trasfondo de conocimientos, no solo va a contrapelo de la cognición humana, sino que además resulta inadecuada para un mundo en el que no todos los hechos relevantes pueden establecerse como premisas.

CATEGORÍAS CLÁSICAS VERSUS CATEGORÍAS DE PARECIDO DE FAMILIA

Una tercera razón por la que la racionalidad jamás se reducirá a la lógica es que los conceptos que importan a la gente difieren de una manera crucial de los predicados de la lógica clásica. Consideremos el predicado «número par», que puede definirse mediante el bicondicional «si un número entero puede dividirse entre dos sin resto, es par, y viceversa». El bicondicional es verdadero, como lo es la proposición «ocho puede dividirse entre dos sin resto», y, a partir de estas premisas verdaderas, podemos deducir la conclusión verdadera «ocho es par». Otro tanto sucede

con «si una persona es mujer y es la madre de un progenitor, es una abuela, y viceversa» y «si una persona es un varón, adulto y no casado, es un soltero, y viceversa». Cabría suponer que, con el suficiente esfuerzo, todos los conceptos humanos pueden definirse de esta manera, estableciendo las condiciones necesarias para que sean verdaderos (el primer *si... entonces* en el bicondicional) y las condiciones suficientes para que sean verdaderos (el converso «viceversa»).

Esta ensoñación fue desbaratada por las célebres objeciones del filósofo Ludwig Wittgenstein (1889-1951).[27] *Intentemos*, decía, hallar las condiciones necesarias y suficientes para cualquiera de nuestros conceptos cotidianos. ¿Cuál es el denominador común de todos los pasatiempos que llamamos *juegos*? ¿La actividad física? No en los juegos de mesa. ¿La alegría? No en el ajedrez. ¿Los competidores? No en el solitario. ¿Ganar y perder? No en «al corro de la patata» ni en un niño lanzando una pelota contra una pared. ¿La habilidad? No en el bingo. ¿El azar? No en los crucigramas. Y Wittgenstein no vivió para ver las artes marciales mixtas, el Pokémon Go o el concurso televisivo *Let's Make a Deal*.[28]

El problema no estriba en que no haya dos juegos que tengan nada en común. Algunos son alegres, como el pillapilla y la charada; algunos tienen ganadores, como el Monopoly y el fútbol; algunos incluyen proyectiles, como el béisbol y la pulga. La tesis de Wittgenstein era que el concepto de *juego* no tiene ningún denominador común que los abarque a todos, ningún conjunto de características necesarias y suficientes que pudiera convertirse en una definición. Antes bien, varios rasgos característicos recorren diferentes subconjuntos de la categoría, de la misma manera que los rasgos físicos pueden encontrarse en diferentes combinaciones en los miembros de una familia. No todo vástago de Robert Kardashian y Kristen Mary Jenner tiene los labios sensuales de las Kardashian ni el pelo negro de las Kardashian ni la piel color caramelo de las Kardashian ni el trasero prominente de las Kardashian. Pero la mayoría de las hermanas tienen algunos de ellos, por lo que podemos reconocer a una Kardashian cuando la vemos, incluso si no existe ninguna proposición verdadera «Si alguien tiene

los rasgos *x*, *y* y *z*, esa persona es una Kardashian». Wittgenstein llegó a la conclusión de que lo que mantiene unidos los miembros de una categoría es el parecido de familia, no las características necesarias y suficientes.

La mayoría de nuestros conceptos cotidianos resultan categorías con parecido de familia, no categorías «clásicas» o «aristotélicas», que son fácilmente estipuladas en la lógica.[29] Estas categorías contienen con frecuencia estereotipos, como el dibujito de un pájaro en un diccionario junto a la definición de *pájaro*, pero la propia definición no acierta a abarcar todos los ejemplares y solo a ellos. La categoría *sillas*, por ejemplo, incluye las sillas de ruedas sin patas, y los taburetes rodantes sin respaldo, los pufs y los accesorios que explotan en las escenas de peleas de Hollywood que no pueden soportar el peso de una persona. Incluso las categorías clásicas que los profesores solían citar para ilustrar el concepto resultan estar plagadas de excepciones. ¿Existe una definición de *madre* que abarque a las madres adoptivas, subrogadas y donantes de óvulos? Si un *soltero* es un hombre no casado, ¿es el papa soltero? ¿Y la mitad masculina de una pareja monógama que nunca se molestó en conseguir el papel del Ayuntamiento? Y hoy en día puedes meterte en muchos problemas si intentas estipular condiciones necesarias y suficientes para ser *mujer*.

Como si esto no bastase para frustrar el sueño de una lógica universal, el hecho de que los conceptos se definan mediante el parecido de familia, más que mediante las condiciones necesarias y suficientes, implica que ni siquiera puede otorgarse a las proposiciones los valores VERDADERO o FALSO. Sus predicados pueden ser más verdaderos con unos sujetos que con otros, dependiendo de lo estereotípico que sea el sujeto, es decir, de cuántos rasgos típicos de la familia posea. Todo el mundo coincide en que «el fútbol es un deporte» es verdadero, pero muchos creen que «la natación sincronizada es un deporte» es, a lo sumo, verídico. Otro tanto cabe decir de «el perejil es una verdura», «una infracción de estacionamiento es un delito», «un derrame cerebral es una enfermedad» y «los escorpiones son bichos». En los juicios cotidianos, la verdad puede ser borrosa.

No es que *todos* los conceptos sean categorías borrosas con parecido de familia.[30] La gente es perfectamente capaz de meter las cosas en cajitas. Todo el mundo entiende que un número es par o impar, sin términos

medios. Bromeamos con que no se puede estar un poquito embarazada o un poquito casado. Comprendemos leyes que evitan disputas interminables sobre casos limítrofes trazando líneas rojas en torno a conceptos tales como *adulto*, *ciudadano*, *propietario*, *esposo* y otras categorías relevantes.

De hecho, una familia entera de falacias informales dimana de las ansias excesivas por pensar en blanco y negro. Tenemos la *falsa dicotomía*: «Naturaleza versus crianza»; «América: ámala o déjala»; «O estás con nosotros o estás con los terroristas»; «O eres parte de la solución o eres parte del problema». Está la falacia de la *pendiente resbaladiza*: «Si legalizamos el aborto, pronto legalizaremos el infanticidio»; «Si permitimos que las personas se casen con un individuo que no es del sexo opuesto, tendremos que permitir que la gente se case con individuos que no son de la misma especie». Y la *paradoja del montón* comienza con la verdad de que si algo es un montón, entonces sigue siendo un montón si quitamos un grano. Pero cuando quitamos otro, y luego otro, llegamos a un punto en el que ya no es un montón, lo cual implica que no existe tal cosa como un montón. Siguiendo la misma lógica, el trabajo se hará incluso si lo pospongo tan solo un día más (la *falacia del mañana*) y no puedo engordar solo por comer una patata frita más (la *falacia de la dieta*).

La respuesta de Wittgenstein a Leibniz y Aristóteles no es solo un tema de debate para los seminarios de filosofía. Muchas de nuestras controversias más feroces implican decisiones acerca de cómo reconciliar los conceptos borrosos basados en el parecido de familia con las categorías clásicas exigidas por la lógica y el derecho. ¿Es un óvulo fertilizado una *persona*? ¿Tuvieron «sexo» Bill y Monica?* ¿Es un vehículo utilitario deportivo un *coche* o una *camioneta*? (Esta última clasificación puso en circulación en Estados Unidos decenas de millones de vehículos que cumplían normas más laxas de seguridad y emisiones.) Y no hace mucho tiempo recibí el siguiente correo electrónico del Partido Demócrata:

> Los republicanos de la Cámara de Representantes están forzando la legislación esta semana para clasificar la *pizza* como «verdura» en las comidas escolares. ¿Por qué? Porque está en marcha una ingente labor de *lobbying* sobre los legisladores republicanos por parte de la industria de la *pizza* congelada [...].

En este Congreso republicano, casi cualquier cosa está en venta para los grupos de presión más poderosos —incluida la definición literal de la palabra *verdura*— y en esta ocasión se está haciendo a expensas de la salud de nuestros hijos.

Firma esta petición y difunde el mensaje: la *pizza* no es verdura.

COMPUTACIÓN LÓGICA VERSUS ASOCIACIÓN DE PATRONES

Si muchos de nuestros juicios son demasiado imprecisos para ser capturados por la lógica, ¿cómo pensamos? Sin las barandillas protectoras de las condiciones necesarias y suficientes, ¿cómo llegamos a ponernos de acuerdo en que el fútbol es un deporte y Kris Jenner es una madre y, a pesar de los republicanos de la Cámara de Representantes, la *pizza* no es una verdura? Si la racionalidad no se implementa en la mente como una lista de proposiciones y una cadena de reglas lógicas, ¿cómo lo hace?

Podemos hallar una respuesta en la familia de modelos cognitivos llamados asociación de patrones, perceptrones, redes conexionistas, modelos de procesamiento distribuido paralelo, redes neuronales artificiales y sistemas de aprendizaje profundo.[31] La idea clave es que, más que manipular cadenas de símbolos con reglas, un sistema inteligente puede agregar decenas, millares o millones de señales graduadas, cada una de las cuales capta el grado en el que está presente una propiedad.

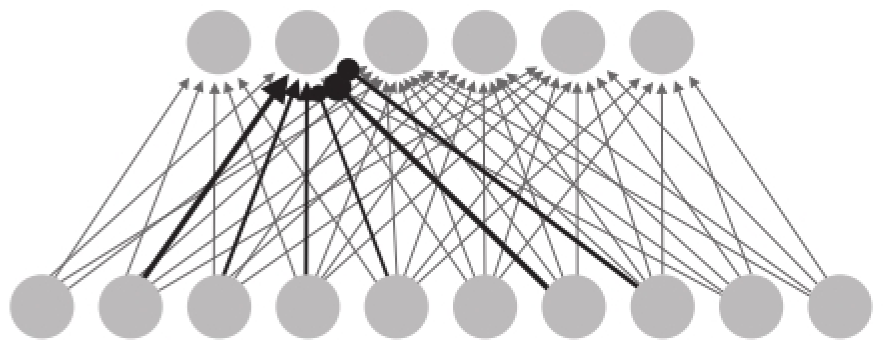
Tomemos el concepto sorprendentemente controvertido de *verdura*. Se trata claramente de una categoría basada en el parecido de familia. No existe ningún taxón linneano que incluya las zanahorias, los brotes de helecho y los champiñones; ningún tipo de órgano de una planta que abarque el brócoli, las espinacas, las patatas, el apio, los guisantes y la berenjena; ni siquiera un gusto, un color o una textura distintivos. Pero, al igual que sucede con las Kardashian, tendemos a conocer las verduras cuando las vemos, porque hay rasgos que se solapan entre los diferentes miembros de la familia. La lechuga es verde, crujiente y de hoja, la espinaca es verde y de hoja, el apio es verde y crujiente, la lombarda es de color morado y de hoja. Cuanto mayor sea el número de características propias de las verduras que tenga algo, y cuanto más definitivamente las posea, más tendemos a llamarlo verdura. La lechuga es una verdura por excelencia; el perejil, no tanto; el ajo, todavía menos. Inversamente, ciertos

rasgos militan en contra de que algo sea una verdura. Aunque algunas verduras rayan en la dulzura, como la calabaza bellota, una vez que una parte de la planta se vuelve demasiado dulce, como un melón cantalupo, lo llamamos, en cambio, fruta. Y, aunque los champiñones de Portobello son carnosos y la calabaza espagueti parece pasta, cualquier cosa hecha de carne animal o masa de harina queda excluida (adiós, *pizza*).

Esto significa que podemos capturar la condición de verdura en una complicada fórmula estadística. Cada una de las características de un ítem (su verdor, su textura crujiente, su dulzura, su pastosidad) se cuantifica, y acto seguido se multiplica por un peso numérico que refleja cuán diagnóstico de la categoría es ese rasgo; positivo alto para el verdor, positivo bajo para la textura crujiente, negativo bajo para la dulzura, negativo alto para la pastosidad. Luego se suman sus valores ponderados y, si la suma excede un determinado umbral, decimos que es una verdura, siendo los mejores ejemplos los números más elevados.

Ahora bien, nadie piensa que formulemos nuestros juicios borrosos llevando a cabo literalmente cadenas de multiplicaciones y adiciones en nuestra cabeza. Pero el equivalente puede realizarse mediante redes de unidades a modo de neuronas, que pueden «activarse» a diferentes ritmos, representando el borroso valor de verdad. Mostramos más abajo una versión de juguete. En la parte inferior tenemos un conjunto de neuronas de entrada alimentadas por los órganos sensoriales, que responden a rasgos simples como *verde* y *crujiente*. En la parte superior tenemos las neuronas de salida, que exhiben la estimación que la red hace de la categoría. Todas las neuronas de entrada están conectadas con todas las neuronas de salida mediante una «sinapsis» de fuerza variable, tanto excitadora (que implementa los multiplicadores positivos) como inhibidora (que implementa los negativos). Las unidades de entrada activadas propagan señales, ponderadas por las fuerzas sinápticas, a las unidades de salida, cada una de las cuales suma el conjunto ponderado de señales entrantes y se activa en consecuencia. En el diagrama, las conexiones excitadoras se representan con flechas, las inhibidoras con puntos, y el grosor de las líneas representa las fuerzas de las sinapsis (mostradas solo para la salida de *verdura*, para simplificar).



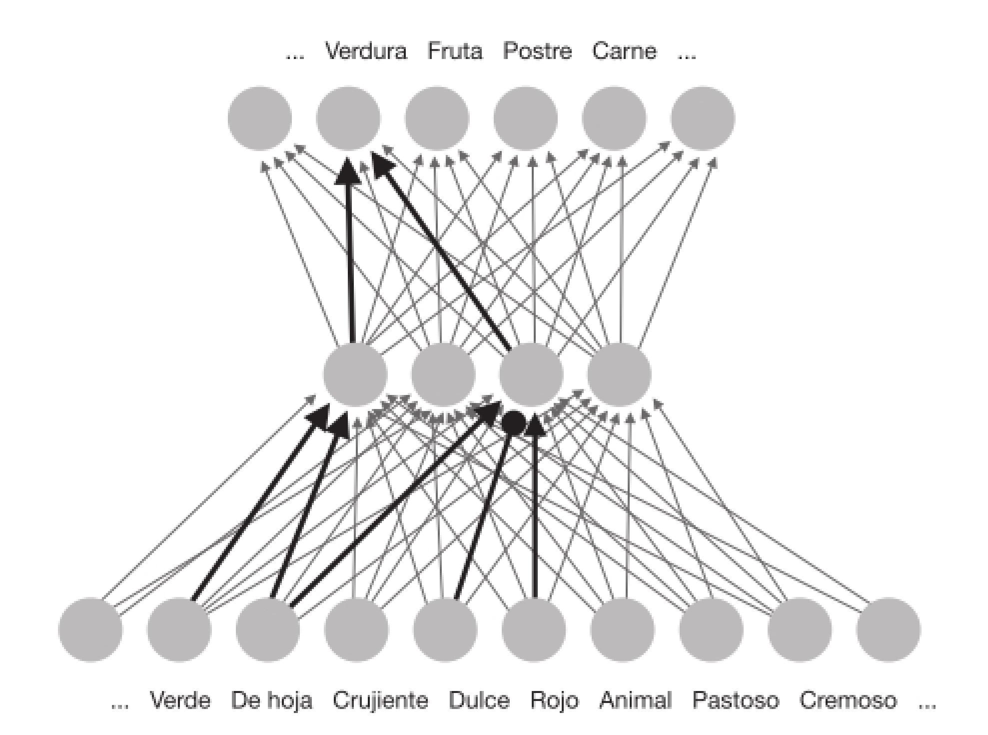


Cabría preguntar quién ha programado esos pesos tan importantes de las conexiones. La respuesta es nadie; se aprenden por experiencia. La red se *entrena* al enfrentarse a muchos ejemplos de diferentes alimentos, junto con la categoría correcta proporcionada por un profesor. La red del neonato, nacida con pequeños pesos aleatorios, ofrece estimaciones débiles y aleatorias. Sin embargo, dispone de un mecanismo de aprendizaje que funciona mediante una regla similar al juego del «frío, frío; caliente, caliente». Compara la salida de cada nodo con el valor correcto ofrecido por el profesor, y sube o baja el peso para cerrar la brecha. Después de cientos de miles de ejemplos de entrenamiento, los pesos de las conexiones se instalan en los mejores valores, y las redes pueden llegar a clasificar muy bien las cosas.

Pero eso solo es cierto cuando las características de entrada indican las categorías de salida de una manera lineal y sumatoria, en virtud de la cual cuanto más, mejor. Funciona para categorías en las que el todo es la suma (ponderada) de sus partes, pero fracasa cuando una categoría se define mediante compensaciones, puntos óptimos, combinaciones ganadoras, píldoras envenenadas, saboteadores de tratos, tormentas perfectas o demasiado de algo bueno. Incluso el simple conector lógico *o* de la disyunción exclusiva, «*x* o *y*, pero no ambos», trasciende los poderes de una red neuronal de dos capas, porque cada una de las características *x* e *y* tiene que potenciar la salida, pero la combinación de ambas tiene que sofocarla. Por consiguiente, aunque una red simple pueda aprender a reconocer las

zanahorias y los gatos, puede fracasar con una categoría rebelde como *verdura*. Un ítem rojo y redondo es probable que sea una fruta si es crujiente y tiene un tallo (como una manzana), pero una verdura si es crujiente y tiene raíces (como una remolacha), o si es carnosa y tiene un tallo (como el tomate). ¿Y qué combinación de colores, formas y texturas podría ser común a los champiñones, las espinacas, la coliflor, las zanahorias y los tomates corazón de buey? Una red de dos capas se confunde con los patrones entrecruzados, tirando del peso hacia arriba y hacia abajo con cada ejemplo de entrenamiento, sin establecerse jamás en valores que separen consistentemente a los miembros de los no miembros.

El problema puede mitigarse insertando una «capa» oculta de neuronas entre la entrada y la salida, como se muestra más abajo. Esto transforma la red de una criatura de estímulo-respuesta en otra con representaciones internas: conceptos, si se quiere. Estas podrían representar aquí categorías cohesivas intermedias como «parecido a la col», «frutas sabrosas», «calabazas», «verduras de hoja», «hongos» y «raíces y tubérculos», cada una de ellas con un conjunto de pesos de entrada que le permiten reconocer el estereotipo correspondiente, y fuertes pesos de salida a *verdura* en la capa de salida.



El reto a la hora de conseguir que estas redes funcionen consiste en cómo entrenarlas. El problema estriba en las conexiones desde la capa de entrada hasta la capa oculta: dado que las unidades están escondidas del entorno, sus conjeturas no pueden cotejarse con los valores «correctos» suministrados por el profesor. Pero un avance en la década de 1980, el algoritmo de aprendizaje por retropropagación de errores, resolvió el problema.[32] En primer lugar, el desajuste entre cada conjetura de la unidad de salida y la respuesta correcta se utiliza para afinar los pesos de las conexiones entre las categorías ocultas y las de salida en la capa superior, al igual que en las redes simples. A continuación, la suma de todos estos errores se propaga hacia atrás hasta cada unidad oculta para ajustar las conexiones entre las categorías de entrada y las ocultas en la capa intermedia. Pareciera que esto no podría funcionar jamás, pero, con millones de ejemplos de entrenamiento, las dos capas de conexiones se instalan en valores que permiten que la red separe el grano de la paja. Igual de extraordinario resulta el hecho de que las unidades ocultas puedan descubrir espontáneamente categorías abstractas como *hongos*, o *raíces y tubérculos*, si eso es lo que las ayuda en la clasificación. Pero lo más frecuente es que las unidades ocultas no representen nada para lo que dispongamos de nombres. Implementan cualesquiera fórmulas complejas que realicen la tarea: «Un poquito de esta característica, pero no demasiado de esa otra, a menos que haya mucho de aquel rasgo».

En la segunda década del siglo XXI, la potencia de los ordenadores se disparó con el desarrollo de las unidades de procesamiento de gráficos, y los datos no dejaban de aumentar a medida que millones de usuarios subían textos e imágenes a la web. Los científicos computacionales lograron suministrar megavitaminas a las redes multicapa, dotándolas de dos, quince, incluso un millar de capas ocultas, y entrenándolas con miles de millones o incluso billones de ejemplos. Estas redes se denominan *sistemas de aprendizaje profundo*, en virtud del número de capas existentes entre la entrada y la salida (no son profundas en el sentido de entender algo). Estas redes están impulsando el «gran despertar de la inteligencia artificial» que estamos viviendo, que está poniendo a nuestra disposición los primeros

productos utilizables para el reconocimiento de voz y de imágenes, la respuesta a preguntas, la traducción y otras hazañas de apariencia humana.[33]

Las redes de aprendizaje profundo con frecuencia tienen un mejor rendimiento que la clásica «inteligencia artificial buena y pasada de moda» (GOFAI, por sus siglas en inglés), que ejecuta deducciones al estilo de la lógica, basadas en proposiciones y reglas codificadas a mano.[34] El contraste en su forma de operar es marcado: a diferencia de la inferencia lógica, el funcionamiento interno de la red neuronal es inescrutable. La mayor parte de los millones de unidades ocultas no representan ningún concepto coherente que podamos comprender, y los científicos computacionales que las entrenan no son capaces de explicar cómo llegan a una respuesta particular. Por eso muchos críticos de la tecnología temen que, a medida que se encomienden a los sistemas de inteligencia artificial decisiones relativas al destino de las personas, estos puedan perpetuar sesgos que nadie pueda identificar y extirpar.[35] En 2018, Henry Kissinger advirtió que, como los sistemas de aprendizaje profundo no se basan en proposiciones que podamos examinar y justificar, auguran el final de la Ilustración.[36] Eso es una exageración, pero el contraste entre la lógica y la computación neuronal es nítido.

¿Es el cerebro humano una gran red de aprendizaje profundo? Ciertamente no, por múltiples razones, pero las similitudes son esclarecedoras. El cerebro tiene en torno a 100.000 millones de neuronas conectadas por 100 billones de sinapsis y, cuando alcanzamos los dieciocho años, hemos estado absorbiendo ejemplos de nuestros entornos durante más de 300 millones de segundos de vigilia. Por consiguiente, estamos preparados para realizar muchas comparaciones y asociaciones de patrones, a la manera de estas redes neuronales. Las redes están hechas a medida para las categorías borrosas basadas en el parecido de familia que constituyen buena parte de nuestro repertorio conceptual. Así pues, las redes neuronales proporcionan pistas sobre la porción de la cognición humana que es racional, pero, técnicamente hablando, no lógica. Desmitifican el inarticulado pero a veces misterioso poder mental que llamamos intuición, instinto, sospechas, corazonadas y sexto sentido.

PESE A LAS COMODIDADES que Siri y Google Translate pueden reportarnos, no debemos pensar que las redes neuronales hayan tornado obsoleta la lógica. Estos sistemas, impulsados por asociaciones borrosas e incapaces de analizar sintácticamente o consultar reglas, pueden ser increíblemente estúpidos.[37] Si le preguntas a Google por «restaurantes de comida rápida cerca de mí que no sean McDonald's», te ofrecerá una lista de todos los McDonald's dentro de un radio de ochenta kilómetros. Pregúntale a Siri: «¿Utilizaba George Washington un ordenador?» y te dirigirá a una reconstrucción por ordenador del rostro de George Washington y a los servicios de sistemas informáticos de la Universidad George Washington. Los módulos de visión que algún día conducirán nuestros coches tienden hoy por hoy a confundir las señales de tráfico con los frigoríficos, y los vehículos volcados con sacos de boxeo, barcos contra incendios y trineos.

La racionalidad humana es un sistema híbrido.[38] El cerebro contiene asociadores de patrones que absorben los parecidos de familia y agregan grandes cantidades de pistas estadísticas. Pero dispone asimismo de un manipulador de símbolos lógicos capaz de ensamblar conceptos en proposiciones y extraer sus implicaciones. Podemos llamarlo sistema 2, cognición recursiva o razonamiento basado en reglas. La lógica formal es una herramienta capaz de depurar y extender este modo de pensar, liberándolo de los errores que acompañan a nuestra condición de animales sociales y emocionales.

Dado que nuestro razonamiento proposicional nos libera de la semejanza y de los estereotipos, posibilita los logros más elevados de la racionalidad humana, tales como la ciencia, la moral y el derecho.[39] Aunque las marsopas guardan parecido de familia con los peces, las reglas que definen la pertenencia a las clases linneanas (como «*si* un animal amamanta a sus crías, *entonces* es un mamífero») nos dicen que de hecho no son miembros. Mediante cadenas de razonamientos categoriales como este, podemos estar convencidos de que los humanos somos simios, el Sol es una estrella y los objetos sólidos son básicamente espacio vacío. En la esfera social, nuestros detectores de patrones ven fácilmente las formas en las que pueden diferir las personas: unos individuos son más ricos, más

inteligentes, más fuertes, más rápidos, más apuestos y más parecidos a nosotros que otros. Pero cuando aceptamos la proposición de que todos los humanos son creados iguales («*si x* es humano, *entonces x* tiene derechos»), podemos aislar estas impresiones de nuestra toma de decisiones legales y morales, y tratar por igual a todas las personas.

CAPÍTULO

4

Probabilidad y aleatoriedad

Mil historias que el ignorante cuenta y cree se desvanecen de repente cuando se ocupa de ellas el informático.

SAMUEL JOHNSON[1]

Aunque Albert Einstein nunca dijo la mayoría de las cosas que supuestamente dijo, sí que afirmó, en diversas versiones: «Jamás creeré que Dios juegue a los dados con el universo».[2] Estuviese o no en lo cierto acerca del mundo subatómico, el mundo en el que vivimos ciertamente *parece* un juego de dados, con impredecibilidad en todas las escalas. La victoria en la carrera no es siempre de los rápidos, ni en la batalla de los fuertes, ni el pan es de los sabios, ni el favor de los hábiles, pero el tiempo y la oportunidad les acontecen a todos. Una parte esencial de la racionalidad consiste en lidiar con la aleatoriedad en nuestras vidas y con la incertidumbre en nuestros conocimientos.

¿Qué es la aleatoriedad? ¿De dónde viene?

En la siguiente tira cómica, la pregunta de Dilbert nos hace tomar conciencia de que la palabra *aleatorio* en lenguaje corriente se refiere a *dos* conceptos: la falta de patrón en los datos y la falta de predictibilidad en un proceso. Cuando duda de que los nueves consecutivos producidos por el trol sean verdaderamente aleatorios, se refiere a su patrón.



La impresión de Dilbert de que hay un patrón en la secuencia no es un producto de su imaginación, como ver mariposas en las manchas de tinta. Los patrones no aleatorios pueden cuantificarse. La brevedad es el alma del

patrón: decimos que un conjunto de datos es no aleatorio cuando su descripción es más breve que el propio conjunto de datos.[3] La descripción «6 9» tiene dos caracteres (en una taquigrafía eficaz para las descripciones), mientras que el propio conjunto de datos «999999» consta de seis caracteres. Otras secuencias que nos parecen no aleatorias también se someten a compresión: «123456» se reduce a «seis primeros»; «505050» se comprime como «tres cincuentas». En contraste, los datos que se nos antojan aleatorios, como «634579», no pueden abreviarse en nada más conciso; han de transcribirse literalmente.

La respuesta del trol captura el segundo sentido de la aleatoriedad: un proceso de generación anárquico e impredecible. El trol tiene razón en que un *proceso* aleatorio puede generar *patrones* no aleatorios, al menos durante un tiempo; en este caso, una salida de seis dígitos. Después de todo, si el generador no tiene ni pies ni cabeza, ¿qué le impide producir seis nueves o cualquier otro patrón no aleatorio, al menos ocasionalmente? Conforme prosigue el generador y la secuencia se vuelve más larga, podemos esperar que se reafirme el patrón aleatorio, porque es improbable que continúe la extraña racha.

El remate del trol es profundo. Como veremos, la confusión de un *patrón* no aleatorio con un *proceso* no aleatorio es uno de los capítulos más extensos de los anales de la locura humana, y el conocimiento de la diferencia entre ellos es uno de los mayores dones de la racionalidad que la educación puede conferir.

Todo esto suscita la pregunta de qué clases de mecanismos físicos pueden generar sucesos aleatorios. A pesar de Einstein, la mayoría de los físicos creen que existe una aleatoriedad irreductible en el reino subatómico de la mecánica cuántica, como la descomposición de un núcleo atómico o la emisión de un fotón cuando un electrón salta de un estado de energía a otro. Es posible que esta incertidumbre cuántica se amplifique a escalas que afecten a nuestras vidas. Cuando yo era ayudante de investigación en un laboratorio de comportamiento animal, las minicomputadoras de la época, del tamaño de un frigorífico, eran demasiado lentas para generar números de aspecto aleatorio en tiempo real, por lo que mi supervisor inventó un artilugio con una cápsula llena de isótopos radiactivos y un contador Geiger

pequeñito que detectaba la pulverización intermitente de partículas y conectaba un interruptor que daba de comer a la paloma.[4] Ahora bien, en la mayor parte del reino de tamaño intermedio en el que transcurren nuestras vidas, los efectos cuánticos se anulan y pueden igualmente no existir.

Así pues, ¿cómo pudo surgir la aleatoriedad en un mundo de bolas de billar que obedecen a las ecuaciones de Newton? Como proclamaba el póster de los años setenta (satirizando los carteles sobre el límite de velocidad): «Gravedad. No es solo una buena idea. Es la ley».[5] El demonio imaginado en 1814 por Pierre-Simon de Laplace, que conocía la posición y el momento lineal de cada partícula del universo, ¿no podría en teoría conectarlos en ecuaciones de las leyes físicas y predecir perfectamente el futuro?

En realidad, existen dos formas en las que un mundo gobernado por leyes puede generar sucesos que, a todos los efectos, sean aleatorios. Una de ellas resultará familiar a los lectores de divulgación científica: el efecto mariposa, cuyo nombre hace referencia a la posibilidad de que el aleteo de las alas de una mariposa en Brasil desencadene un tornado en Texas. Los efectos mariposa pueden producirse en sistemas dinámicos deterministas no lineales, también conocidos como *caos*, en los que minúsculas diferencias en las condiciones iniciales, demasiado pequeñas para ser medidas por cualquier instrumento, pueden alimentarse a sí mismas y estallar en efectos colosales.

La otra manera en la que un sistema determinista puede parecer aleatorio desde un punto de vista humano tiene también un nombre familiar: *cara o cruz*. El destino de una moneda lanzada al aire no es literalmente aleatorio; un mago experto puede lanzar una de tal manera que saque una cara o una cruz a demanda. Pero, cuando un resultado depende de un gran número de pequeñas causas a las que resulta poco factible realizar un seguimiento, como los ángulos y la fuerza con los que se lanza la moneda y las corrientes de aire que la zarandean, también puede ser aleatorio.

¿QUÉ SIGNIFICA *PROBABILIDAD*?

Cuando el meteorólogo de televisión dice que hay un 30 % de probabilidad de lluvia en la región mañana, ¿qué quiere decir? Para la mayoría de la gente, la respuesta es nebulosa. Algunos creen que significa que lloverá en el 30 % de la región. Otros piensan que significa que lloverá el 30 % del tiempo. Unos pocos piensan que significa que el 30 % de los meteorólogos creen que lloverá. Y algunos piensan que significa que lloverá en algún lugar de la región el 30 % de los días en los que se hace una predicción como esa (esto último es, de hecho, lo que más se aproxima a lo que el meteorólogo tenía en mente).[6]

Los observadores climáticos no son los únicos que se sienten confusos. En 1929, Bertrand Russell observaba que el de «*probabilidad* es el concepto más importante de la ciencia moderna, especialmente porque nadie tiene la más remota idea de lo que significa».[7] Para ser más exactos, diferentes personas tienen ideas distintas de lo que significa, como vimos en el capítulo 1 con los problemas de Monty Hall y Linda.[8]

Tenemos la definición *clásica* de probabilidad, que se remonta a los orígenes de la teoría de la probabilidad como una forma de entender los juegos de azar. Presentamos los posibles resultados de un proceso que tiene una igual probabilidad de ocurrir, sumamos los que cuentan como ejemplos del suceso y dividimos entre el número de posibilidades. Un dado puede aterrizar en cualquiera de las seis caras. Un «número par» corresponde a su aterrizaje en las caras con dos, cuatro o seis puntos. De un total de seis posibilidades, puede aterrizar en «par» de tres formas. Decimos que la probabilidad clásica de que salga «par» es de tres de seis, o 0,5 (en el capítulo 1 utilicé la definición clásica para explicar la estrategia correcta en el dilema de Monty Hall y observé que los errores en el cálculo de los resultados posibles eran los que llevaban a algunos de los expertos demasiado confiados a adoptar la estrategia incorrecta).

Ahora bien, ¿por qué pensábamos de entrada que el resultado del aterrizaje en cada cara tenía la misma probabilidad de ocurrir? Evaluábamos la *propensión* del dado, su disposición física a hacer varias cosas. Esta incluye la simetría de las seis caras, la manera azarosa en que lanza el dado el tirador y la física de la caída.

Íntimamente relacionada está una tercera interpretación *subjetivista*. Antes de lanzar el dado, basándote en todo lo que sabes, ¿cómo cuantificarías, en una escala de cero a uno, tu creencia en que saldrá par? Esta estimación de la creencia se denomina a veces *interpretación bayesiana de la probabilidad* (un tanto engañosamente, como veremos en el próximo capítulo).

Luego está la *interpretación evidencial*: el grado en el que crees que la información presentada garantiza la conclusión. Pensemos en un tribunal de justicia donde, al juzgar la probabilidad de que el acusado sea culpable, ignoramos la información general inadmisible y prejuiciosa, y consideramos únicamente la fuerza de los argumentos expuestos por el fiscal. Era la interpretación basada en las evidencias la que hacía racional juzgar que Linda, habiendo sido presentada como una luchadora por la justicia social, tuviese más probabilidades de ser una cajera feminista que una cajera.

Finalmente, está la *interpretación frecuentista*: si lanzásemos efectivamente el dado muchas veces, pongamos por caso mil, y contáramos los resultados, descubriríamos que saldría un número par aproximadamente en quinientos lanzamientos, la mitad de ellos.

Ordinariamente las cinco interpretaciones están alineadas. En el caso de un lanzamiento de moneda, esta es simétrica; las caras representan exactamente uno de los dos resultados posibles; tu corazonada está a mitad de camino entre «seguro que sale cara» y «seguro que sale cruz»; el argumento a favor de la cara es tan sólido como el argumento a favor de la cruz; y, a la larga, la mitad de los lanzamientos que verás serán caras. La probabilidad de que salga cara es 0,5 en todos los casos. Pero las interpretaciones no significan lo mismo y a veces se separan. En tales casos, los enunciados acerca de las probabilidades pueden generar confusión, controversia e incluso tragedias.

Lo más dramático es que las cuatro primeras interpretaciones son aplicables a la noción vagamente mística de la probabilidad de un único caso. ¿Cuál es la probabilidad de que tengas más de cincuenta años? ¿De que el próximo papa sea Bono? ¿De que Britney Spears y Katy Perry sean la misma persona? ¿De que exista vida en Encélado, una de las lunas de

Saturno? Cabría objetar que estas preguntas carecen de sentido: o bien tienes más de cincuenta años o bien no los tienes, y la probabilidad no tiene nada que ver con ello. No obstante, en la interpretación subjetivista puedo poner un número a mi ignorancia. Esto ofende a algunos estadísticos, que quieren reservar el concepto de *probabilidad* para la frecuencia relativa en un conjunto de sucesos, que son verdaderamente reales y pueden contarse. Uno de ellos bromeaba diciendo que las probabilidades de los sucesos únicos no pertenecen a las matemáticas, sino al psicoanálisis.[9]

También los legos pueden tener dificultades para comprender el concepto de *probabilidad numérica* de un solo acontecimiento. Se enojan con el hombre del tiempo al empaparse un día que este había pronosticado una probabilidad de lluvia del 10 %, y se ríen del agregador de encuestas que predijo que Hillary Clinton tenía un probabilidad del 60 % de ganar las elecciones presidenciales de 2016. Pero estos adivinos se defienden invocando una interpretación frecuentista de sus probabilidades: uno de cada diez días en los que hace una predicción semejante, llueve; en seis de cada diez elecciones con esas cifras en las encuestas, gana el candidato principal. En esta tira cómica, el jefe de Dilbert ilustra una falacia común:



Como vimos en el capítulo 1 con Linda y veremos de nuevo en el próximo capítulo, la reformulación de la probabilidad de la creencia en un único suceso como frecuencia en un conjunto de sucesos puede recalibrar las intuiciones de las personas. Un fiscal en una gran ciudad que diga: «La probabilidad de que el ADN en la ropa de la víctima coincida con el ADN del sospechoso si este fuera inocente es de uno de cien mil» tiene más probabilidades de conseguir una condena que otro que diga: «De cada cien mil personas inocentes de esta ciudad, una mostrará una coincidencia». Lo primero parece una estimación de la duda subjetiva que es indistinguible de cero; lo segundo nos invita a imaginar a ese tipo acusado falsamente, junto con los muchos otros que viven en la metrópoli.

La gente confunde asimismo la probabilidad en el sentido frecuentista con la propensión. Gerd Gigerenzer cuenta una visita a una fábrica aeroespacial en la que el guía explicó a los visitantes que sus cohetes Ariane tenían un factor de seguridad del 99,6 %.[10] Estaban delante de un cartel que representaba los noventa y cuatro cohetes y sus historias respectivas, y ocho de ellos se habían estrellado o explotaron. Cuando Gigerenzer preguntó cómo un cohete con un factor de seguridad del 99,6 % podía fallar casi el 9 % de las veces, el guía explicó que el factor se calculaba a partir de las fiabilidades de las piezas individuales, y los fallos eran el resultado de errores humanos. Por supuesto, lo que nos importa a la postre es con cuánta frecuencia escapa en total el cohete de los lazos opresores de la tierra[*] o se destruye, independientemente de las causas, por lo que la única probabilidad relevante es la frecuencia total. En virtud del mismo malentendido, la gente se pregunta a veces por qué a un candidato popular que está muy por delante en las encuestas se le otorga solamente una probabilidad del 60 % de ganar las elecciones, cuando solo una sorpresa de última hora podría hacerle descarrilar. La respuesta es que la estimación de la probabilidad tiene en cuenta las sorpresas de última hora.

PROBABILIDAD VERSUS DISPONIBILIDAD

Pese a la diferencia de interpretaciones, la probabilidad está íntimamente vinculada a los sucesos como una proporción de las oportunidades, ya sea directamente, en las definiciones clásica y frecuentista, ya indirectamente, con los otros juicios. Sin duda, cada vez que decimos que un suceso es más probable que otro, creemos que ocurrirá con más frecuencia, dada la oportunidad. Para calcular el riesgo, deberíamos sumar el número de casos de un acontecimiento y dividirlo mentalmente entre el número de ocasiones en que este podría haber tenido lugar.

No obstante, uno de los hallazgos característicos de la ciencia del juicio humano es que no es así como funciona generalmente la estimación de la probabilidad. Antes bien, las personas juzgan la probabilidad de los sucesos por la facilidad con la que los casos vienen a la mente, un hábito que Tversky y Kahneman denominaron *heurística de la disponibilidad*.[11] Utilizamos la clasificación del motor de búsqueda de nuestro cerebro (imágenes, anécdotas y vídeos mentales que este arroja) como nuestra mejor estimación de las probabilidades. La heurística explota una característica de la memoria humana, a saber: que el recuerdo se ve afectado por la frecuencia; cuanto más a menudo encontramos algo, más fuerte es el rastro que deja en nuestro cerebro. Por lo tanto, trabajar hacia atrás y calcular la frecuencia partiendo de la recordabilidad a menudo funciona bien en la práctica. Cuando te piden que adivines cuáles son las aves más comunes en una ciudad, no harías mal en apostar por las palomas y los gorriones más que por los ampelis y los papamoscas, en lugar de tomarte la molestia de consultar un censo de aves.

Durante la mayor parte de la existencia humana, la disponibilidad y las habladurías eran las *únicas* formas de calcular la frecuencia. Algunos Gobiernos disponían de bases de datos estadísticas, pero estas se consideraban secretos de Estado y únicamente se divulgaban a las élites administrativas. Con el surgimiento de las democracias liberales en el siglo XIX, los datos pasaron a considerarse un bien público.[12] Incluso hoy, cuando los datos acerca de prácticamente todo están a unos cuantos clics, no son muchos los que sacan partido de ellos. Recurrimos instintivamente a nuestras impresiones, que distorsionan nuestra comprensión cuando la fortaleza de esas impresiones no refleja las frecuencias en el mundo. Eso

puede suceder cuando nuestras experiencias son una muestra sesgada de esos sucesos, o cuando las impresiones son promovidas o degradadas en nuestros resultados de búsquedas mentales por amplificadores psicológicos tales como la recencia, la vivacidad o la intensidad emocional. Los efectos sobre los asuntos humanos son arrolladores.

Al margen de nuestra experiencia inmediata, aprendemos acerca del mundo a través de los medios de comunicación. La cobertura mediática impulsa así la sensación de frecuencia y de riesgo de los individuos: estos creen que tienen más probabilidades de morir por un tornado que por asma, aunque el asma sea ochenta veces más letal, presumiblemente porque los tornados son más fotogénicos.[13] Por razones similares, las clases de personas que no pueden mantenerse al margen de las noticias tienden a estar sobrerrepresentadas en nuestros censos mentales. ¿Qué porcentaje de chicas adolescentes dan a luz cada año a escala mundial? La gente supone que el 20 %, en torno a diez veces el porcentaje real. ¿Qué proporción de estadounidenses son inmigrantes? Alrededor del 28 %, dicen los encuestados; la respuesta correcta es el 12 %. ¿Homosexuales? Los estadounidenses suponen que el 24 %; las encuestas indican que el 4,5 %.[14] ¿Afroamericanos? En torno a un tercio, dice la gente, unas dos veces y media por encima de la cifra real, 12,7 %. Esto sigue siendo más exacto que su estimación de otra minoría conspicua, los judíos, con quienes los encuestados se desvían en un factor de nueve (el 18 % frente al 2 %).[15]

La heurística de la disponibilidad es un motor de acontecimientos mundiales, con frecuencia en direcciones irracionales. Aparte de la enfermedad, el riesgo más letal para la vida y la integridad física son los accidentes, que matan alrededor de cinco millones de personas al año (de un total de 56 millones de muertes), en torno a una cuarta parte de ellos en accidentes de tráfico.[16] No obstante, excepto cuando se cobran la vida de una celebridad fotogénica, los accidentes de coche rara vez aparecen en las noticias y la gente se despreocupa de la matanza. En cambio, los accidentes de aviación consiguen una generosa cobertura, pero solamente matan a unas doscientas cincuenta personas al año en el mundo entero, por lo que los aviones son alrededor de mil veces más seguros por pasajero y milla recorrida que los coches.[17] Sin embargo, todos conocemos a personas con

miedo a volar, pero a nadie con miedo a la conducción, y un sangriento accidente de avión puede asustar a los pasajeros aéreos y lanzarlos a las carreteras durante los meses siguientes, donde morirán miles más.[18] La viñeta de *SMBC* hace hincapié en algo similar.

Entre las muertes más vívidas y espantosas imaginables figura la descrita en una canción de *La ópera de los tres centavos*: «Cuando ese tiburón muerde con sus dientes, cariño, comienzan a esparcirse las olas escarlata».[19*] En 2019, cuando un surfista de Cape Cod se convirtió en la primera víctima mortal en Massachusetts en más de ochenta años, las localidades equiparon todas las playas con amenazadores carteles de advertencia con forma de mandíbulas y kits de control de hemorragias, y encargaron estudios sobre torres, drones, aviones, globos, sonares, boyas acústicas y repelentes electromagnéticos olorosos. Sin embargo, cada año mueren en Cape Cod entre quince y veinte personas en accidentes de tráfico, y las mejoras baratas en señalización, barreras y aplicación de las normas de tráfico podrían salvar muchas más vidas con una fracción de los costes.[20]



Utilizado con el permiso de Zach Weinersmith

El sesgo de disponibilidad puede afectar al destino del planeta. Después de hacer números, varios científicos climáticos eminentes advierten que «no existe ningún camino creíble hacia la estabilización climática que no otorgue un papel sustancial a la energía nuclear».[21] La energía nuclear es la forma más segura de energía jamás utilizada por la humanidad. En los accidentes mineros, los fallos en las centrales hidroeléctricas, las explosiones de gas natural y los accidentes de trenes petroleros muere gente, a veces en grandes cantidades, y el humo de la combustión de carbón provoca un número enorme de muertes, más de medio millón al año. Sin embargo, la energía nuclear lleva décadas estancada en Estados Unidos y está retrocediendo en Europa, reemplazada con frecuencia por el sucio y peligroso carbón. En gran parte, la oposición es espoleada por los recuerdos de tres accidentes: Three Mile Island en 1979, que no mató a nadie; Fukushima en 2011, que mató a un operario

años después (las otras muertes fueron causadas por el tsunami y por el pánico de la evacuación); y la chapuza soviética de Chernóbil en 1986, que mató a treinta y una personas en el accidente y quizá a varios miles de cáncer, aproximadamente el mismo número de muertes por emisiones de carbón *cada día*.[22]

Sin lugar a dudas, la disponibilidad no es el único distorsionador de la percepción del riesgo. Paul Slovic, un colaborador de Tversky y Kahneman, demostró que los individuos sobrestiman asimismo el peligro de las amenazas que son nuevas (más vale lo malo conocido que lo bueno por conocer), están fuera de su control (como si pudieran conducir con más seguridad que con la que puede volar un piloto), son creadas por los humanos (de modo que evitan los alimentos genéticamente modificados, pero se tragan las muchas toxinas que han evolucionado naturalmente en las plantas) y perciben como no equitativas (cuando sienten que asumen un riesgo en beneficio de otro).[23] Cuando estas pesadillas se combinan con la perspectiva de un desastre que mate a muchas personas a la vez, la suma de todos los temores se convierte en un *riesgo aterrador*. Los accidentes de aviación, los accidentes nucleares y los ataques terroristas son los mejores ejemplos.

EL TERRORISMO, COMO OTRAS PÉRDIDAS de vidas con premeditación, genera una química del miedo diferente. Los científicos de datos que hacen recuentos de víctimas se quedan perplejos con frecuencia ante la forma en que las matanzas muy publicitadas, pero con pocas víctimas, pueden provocar reacciones sociales que hacen época. El 11S fue con mucho el peor atentado terrorista de la historia y se cobró tres mil vidas; en la mayor parte de los años malos, Estados Unidos sufre unas pocas docenas de muertes por terrorismo, un error de redondeo en el recuento de homicidios y accidentes (la cifra anual es más baja, por ejemplo, que el número de personas muertas por rayos, picaduras de abejas o ahogamientos en bañeras). Con todo, el 11S condujo a la creación de un nuevo departamento federal, a la vigilancia masiva de los ciudadanos y al endurecimiento de los

servicios públicos, así como a dos guerras que mataron a más del doble de los estadounidenses que murieron en 2001, junto con centenares de miles de iraquíes y afganos.[24]

Por mencionar otro peligro aterrador con pocas muertes, las masacres en los colegios estadounidenses se cobran unas treinta y cinco víctimas al año, comparadas con alrededor de los dieciséis mil homicidios en los registros policiales rutinarios.[25] Sin embargo, las escuelas estadounidenses han implementado dudosas medidas de seguridad que han costado miles de millones de dólares, como instalar pizarras blancas a prueba de balas y armar a los profesores con pistolas no letales, traumatizando al mismo tiempo a los niños con aterradores simulacros de tiroteo activo. En 2020, el brutal asesinato de George Floyd, un afroamericano desarmado, a manos de un agente de policía blanco, provocó protestas masivas y la adopción súbita de una doctrina académica radical, la teoría crítica de la raza, por parte de universidades, periódicos y corporaciones. Estas convulsiones eran atizadas por la impresión de que los afroamericanos corren un grave peligro de morir a manos de la policía. No obstante, como sucede con el terrorismo y los tiroteos en las escuelas, las cifras son sorprendentes. Un total de sesenta y cinco estadounidenses desarmados de todas las razas son asesinados por la policía en un año medio, de los cuales veintitrés son afroamericanos, lo que representa en torno a tres décimas del 1 % de las siete mil quinientas víctimas de homicidios afroamericanas.[26]

Resultaría psicológicamente obtuso explicar la reacción desproporcionada a las matanzas publicitadas exclusivamente por el miedo exacerbado por la disponibilidad. Como ocurre con muchos signos de aparente irracionalidad, intervienen aquí otras lógicas al servicio de objetivos distintos de las probabilidades precisas.

Nuestra reacción desproporcionada al asesinato más monstruoso puede resultar irracional en el marco de la teoría de la probabilidad, pero racional en el marco de la teoría de juegos (capítulo 8). El homicidio no es como otros peligros mortales. A un huracán o a un tiburón no le importa cómo responderemos al daño que tienen reservado para nosotros, pero puede que a un asesino humano sí. Por consiguiente, cuando la gente reacciona a un asesinato con conmoción e indignación públicas, y redobla su compromiso

con la autodefensa, la justicia o la venganza, envía una señal a los asesinos potenciales de ahí afuera, posiblemente haciendo que se lo piensen dos veces.

La teoría de juegos puede explicar asimismo el frenesí desencadenado por una clase especial de suceso que Thomas Schelling describió en 1960, que cabe denominar ultraje comunitario.[27] Un ultraje comunitario es un ataque flagrante y ampliamente atestiguado contra un miembro o símbolo de un colectivo. Se percibe como una afrenta intolerable e incita a la colectividad a sublevarse y vengarse para hacer justicia. Entre los ejemplos se incluyen la explosión del USS Maine en 1889, que condujo a la guerra hispano-estadounidense; el hundimiento del RMS Lusitania en 1915, que desencadenó la entrada de Estados Unidos en la Primera Guerra Mundial; el incendio del Reichstag de 1933, que posibilitó la instauración del régimen nazi; Pearl Harbor en 1941, que envió a América a la Segunda Guerra Mundial; el 11S, que desembocó en las invasiones de Afganistán e Irak; y el hostigamiento a un vendedor ambulante en Túnez en 2010, cuya autoinmolación desató la Revolución tunecina y la Primavera Árabe. La lógica de estas reacciones es el *conocimiento común*, en el sentido técnico de algo que todos saben que todos saben que todos saben.[28] El conocimiento común es necesario para la *coordinación*, en la que varias partes actúan con la expectativa de que cada una de las demás también lo hará. El conocimiento común puede ser generado por *puntos focales*, acontecimientos públicos que las personas ven que otras personas ven. Un ultraje público puede ser el conocimiento común que soluciona el problema de lograr que todos actúen en concierto cuando la vejación ha ido creciendo gradualmente y nunca parece llegar el momento oportuno para reaccionar ante ella. Un ultraje que no se puede ignorar puede desencadenar la indignación simultánea en una población dispersa y convertirla en un colectivo resuelto. La cantidad de daño infligido por el ataque es irrelevante.

No solo irrelevante, sino también tabú. Un ultraje comunitario inspira lo que el psicólogo Roy Baumeister denomina una *narrativa de las víctimas*: una alegoría moralizada en la que se santifica un acto perjudicial y el daño se consagra como irreparable e imperdonable.[29] El objetivo de la

narrativa no es la exactitud, sino la solidaridad. La puntillosidad respecto de lo realmente acontecido no solo es irrelevante, sino también sacrílega o traidora.[30]

En el mejor de los casos, un ultraje público puede movilizar la acción postergada contra un problema de larga data, como está sucediendo en la lucha contra el racismo sistémico en respuesta al asesinato de Floyd. Un liderazgo reflexivo puede canalizar un ultraje hacia una reforma responsable, como se plasma en el lema político «Nunca dejes que una crisis se desperdicie».[31] Pero la historia de los ultrajes públicos sugiere que estos también pueden empoderar a los demagogos y abocar a las turbas hacia los cenagales y los desastres. En conjunto, sospecho que son mayores los beneficios obtenidos cuando las cabezas más frías evalúan los daños con precisión y responden a ellos de manera proporcionada.[32]

LOS ULTRAJES NO PUEDEN DEVENIR públicos sin cobertura mediática. Fue a raíz de la explosión del Maine cuando el término *prensa amarilla* llegó a ser moneda corriente. Incluso cuando los periodistas no provocan en los lectores una histeria patriotera, las reacciones públicas desmedidas constituyen un peligro inherente. A mi juicio, los periodistas no han reflexionado lo suficiente sobre el modo en que la cobertura mediática puede activar nuestros sesgos cognitivos y distorsionar nuestra comprensión. Los cínicos podrían responder que a los periodistas eso les importa un comino, pues lo único que les preo cupa son los clics y los globos oculares. No obstante, por mi experiencia, la mayoría de los periodistas son idealistas que creen responder a la más elevada vocación de informar al público.

La prensa es una máquina de disponibilidad. Ofrece anécdotas que alimentan nuestra impresión de lo que es común de tal manera que está garantizado que será engañosa. Dado que la noticia es lo que sucede, no lo que no sucede, el denominador de la fracción correspondiente a la probabilidad de la verdad de un acontecimiento (todas las oportunidades de que se produzca el suceso, incluidas aquellas en las que no ocurre) es invisible, dejándonos en la oscuridad acerca de la auténtica prevalencia de algo.

Por otra parte, las distorsiones no son azarosas, sino que nos desvían hacia lo morboso. Las cosas que suceden de repente suelen ser malas (una guerra, un tiroteo, una hambruna, un colapso financiero), pero las cosas buenas pueden consistir en que no suceda nada, como un aburrido país en paz o una región poco memorable que goza de salud y buena alimentación. Y cuando se producen progresos, estos no se forjan en un día, sino que crecen unos pocos puntos porcentuales al año, transformando el mundo con sigilo. Como señala el economista Max Roser, los sitios web de noticias podrían haber publicado el titular «137.000 personas escaparon ayer de la pobreza extrema» todos los días durante los últimos veinticinco años.[33] Pero nunca publicaron ese titular, porque nunca hubo un jueves de octubre en el que eso sucediera de repente. Por consiguiente, uno de los mayores progresos de la historia humana, la escapatoria de la miseria de 1.250 millones de personas, ha pasado desapercibido.

La ignorancia es mensurable. Los encuestadores descubren reiteradamente que, mientras que las personas tienden a ser demasiado optimistas en lo que atañe a sus propias vidas, son demasiado pesimistas con respecto a sus sociedades. Por ejemplo, en la mayoría de los años entre 1992 y 2015, una época que los criminólogos describen como la gran caída del crimen en Estados Unidos, una mayoría de estadounidenses creían que los crímenes estaban aumentando.[34] En su Proyecto sobre la Ignorancia, Hans y Ola Rosling y Anna RoslingRönnlund han demostrado que la comprensión de las tendencias globales en la mayoría de las personas educadas es justamente al revés: piensan que la longevidad, la alfabetización y la pobreza extrema están empeorando, mientras que todas ellas han mejorado espectacularmente[35] (la pandemia de la COVID-19 invirtió estas tendencias en 2020, casi con certeza de manera temporal).

La ignorancia impulsada por la disponibilidad puede resultar corrosiva. Un noticiario mental en bucle de las catástrofes y los fracasos puede alimentar el cinismo acerca de la capacidad de la ciencia, la democracia liberal y las instituciones de cooperación global para mejorar la condición humana. El resultado puede ser un fatalismo paralizante o un radicalismo temerario: una llamada a destruir la máquina, drenar el pantano o empoderar a un demagogo que prometa: «Yo soy el único que puede

solucionarlo».[36] El periodismo que trafica con las calamidades establece asimismo perversos incentivos para los terroristas y los tiradores desbocados, que pueden explotar el sistema y ganar notoriedad instantánea.[37] Y un lugar especial en el infierno de los periodistas está reservado para los escribas que en 2021, durante el despliegue de las vacunas contra la COVID, que sabemos que tienen una tasa de eficacia del 95 %, escribieron historias sobre las personas vacunadas que contrajeron la enfermedad, lo cual por definición no es noticia (pues siempre se supo con seguridad que habría algunas) y ahuyentó a miles de individuos de un fármaco que salva vidas.

¿Cómo podemos reconocer los peligros genuinos en el mundo al tiempo que calibramos nuestra comprensión en sintonía con la realidad? Los consumidores de noticias deberían ser conscientes de su sesgo inherente y adaptar su dieta informativa para incluir fuentes que presenten un panorama estadístico más amplio: menos canales de noticias de Facebook y más *Our World in Data [Nuestro mundo en datos]*.[38] Los periodistas deberían poner los sucesos escabrosos en contexto. Una matanza, un accidente de aviación o el ataque de un tiburón deberían ir acompañados de la tasa anual, que tenga en cuenta el denominador de la probabilidad, no solo el numerador. Un revés o una avalancha de desgracias deberían compararse con tendencias a más largo plazo. Las fuentes informativas podrían incluir un tablero de indicadores nacionales y mundiales (tasa de homicidios, emisiones de dióxido de carbono, muertos de guerra, democracias, delitos de odio, violencia contra las mujeres, pobreza y así sucesivamente), de suerte que los lectores puedan ver las tendencias por sí mismos y hacerse una idea de qué políticas mueven la aguja en la dirección adecuada. Aunque los editores me han contado que los lectores odian las matemáticas y jamás tolerarán que los números estropeen sus historias y sus fotografías, sus propios medios de comunicación desmienten esta condescendencia. La gente consume ávidamente datos en las páginas dedicadas al tiempo, a los negocios y a los deportes, ¿por qué no entonces en las noticias?

Probabilidades conjuntivas, disyuntivas y condicionales

Un meteorólogo de televisión anuncia que hay un 50 % de probabilidades de lluvia el sábado y un 50 % de probabilidades de lluvia el domingo, y concluye que hay un cien por cien de probabilidades de lluvia durante el fin de semana.[39] En un viejo chiste, un hombre lleva consigo una bomba en un avión por su propia seguridad, porque, piensa, ¿cuáles son las probabilidades de que haya *dos* bombas en un avión? Y luego está el argumento de que el papa es casi con certeza un alienígena. La probabilidad de que una persona seleccionada al azar en la Tierra sea el papa es ínfima: uno entre 7.800 millones, o 0,00000000013. Francisco es el papa. Por tanto, probablemente Francisco no sea un ser humano.[40]

Al razonar acerca de la probabilidad, resulta fácil descarrilarse. Estas meteduras de pata derivan de la aplicación errónea del siguiente paso en la comprensión de la probabilidad: cómo calcular las probabilidades de una conjunción, una disyunción, un complemento y un condicional. Si estos términos suenan familiares, es porque son simplemente los equivalentes probabilísticos de *y*, *o*, *no* y *si... entonces* del capítulo anterior. Aunque las fórmulas son sencillas, cada una de ellas tiende una trampa, y los tropiezos son los que dan lugar a las pifias con la probabilidad.[41]

La probabilidad de una conjunción de dos eventos independientes, P(A Y B), es el producto de las probabilidades de cada uno de ellos: P(A) × P(B). Si los Green tienen dos hijos, ¿cuál es la probabilidad de que ambos sean niñas? Es la probabilidad de que el primero sea una niña, 0,5 veces, y la probabilidad de que el segundo sea una niña, también 0,5, o 0,25. Traduciendo del lenguaje del suceso único al frecuentista, nos encontraremos con que, del total de familias con dos hijos que consideremos, una cuarta parte tendrán dos niñas. Más intuitivamente aún, la definición clásica de probabilidad nos aconseja plantear las posibilidades lógicas: niño-niño, niño-niña, niña-niño, niña-niña. Una de estas cuatro es las dos niñas.

La trampa de la fórmula de la conjunción reside en la condición de *independientes*. Los sucesos son independientes cuando están desconectados: la probabilidad de ver uno de ellos no influye en la

probabilidad de ver el otro. Imaginemos una sociedad, tal vez no remota, en la que las personas puedan elegir el sexo de sus hijos. A título de ejemplo, imaginemos que los padres son sexistas en cuestión de género, de suerte que la mitad quiere solo niños y la otra mitad solo niñas. Si el primer hijo es una niña, nos avisa de que los padres preferían una niña, lo cual significa que probablemente volverían a optar por una niña, y viceversa si el primer hijo es un niño. Los sucesos no son independientes y la multiplicación falla. Si las preferencias fueran absolutas y la tecnología perfecta, cada familia tendría solo niños o solo niñas, y la probabilidad de que en una familia de dos hijos ambos sean niñas sería 0,5, no 0,25.

No pensar si los sucesos son independientes puede conducir a grandes meteduras de pata. Cuando aparece una racha de casos poco frecuentes en entidades que no están en cuarentena las unas respecto de las otras —los ocupantes de un edificio que se contagian el resfriado unos a otros, o los miembros de un grupo de pares que se copian sus modas respectivas, o las respuestas de la encuesta de un único participante que mantiene sus sesgos de una pregunta a otra, o las mediciones de algo en días, meses o años sucesivos, que pueden mostrar inercia—, entonces el conjunto de observaciones es en efecto un suceso único, no un curso caprichoso de acontecimientos, y sus probabilidades no se pueden multiplicar. Por ejemplo, si el índice de criminalidad se situase por debajo de la media en cada uno de los doce meses después de colocar carteles de vigilancia vecinal en una ciudad, sería un error concluir que la racha debe responder a los letreros en lugar de al azar. Los índices de criminalidad varían lentamente, arrastrándose los patrones de un mes al siguiente, de suerte que el resultado se aproxima más a un único lanzamiento de moneda que a una serie de doce lanzamientos.

En el ámbito jurídico, la aplicación errónea de la fórmula para una conjunción no es solo un error matemático, sino también un error judicial. Un ejemplo notorio es la falaz ley de Meadow, que lleva el nombre de un pediatra británico que declaró que, al examinar las muertes súbitas de lactantes en una familia, «una es una tragedia, dos son sospechosas y tres son un asesinato, a menos que exista una prueba en contra». En el caso de 1999 de la abogada Sally Clark, que había perdido dos bebés varones, el

médico testificó que, como la probabilidad de una muerte en la cuna en una familia acomodada y no fumadora es de una entre 8.500, la probabilidad de dos muertes en la cuna es el cuadrado de ese número, uno entre 72 millones. Clark fue condenada a cadena perpetua por asesinato. Los estadísticos horrorizados señalaron el error: las muertes súbitas de lactantes en el seno de una familia no son independientes, ya que los hermanos pueden compartir una predisposición genética, el hogar puede tener factores de riesgo elevados y los padres pueden haber reaccionado a la primera tragedia tomando precauciones equivocadas que incrementaran la probabilidad de un segundo caso. Clark fue puesta en libertad tras una segunda apelación (por diferentes motivos) y en los años siguientes tuvieron que ser revisados centenares de casos basados en errores similares.[42]

Otro error garrafal en el cálculo de las conjunciones hizo su aparición estelar en un estrafalario intento de Donald Trump y sus seguidores de invertir los resultados de las elecciones presidenciales de 2020 basándose en acusaciones infundadas de fraude electoral. En una moción presentada ante la Corte Suprema de Estados Unidos, el fiscal general de Texas Ken Paxton escribió: «La probabilidad de que el exvicepresidente Biden gane el voto popular en los cuatro estados demandados (Georgia, Míchigan, Pensilvania y Wisconsin) independientemente, dada la ventaja inicial del presidente Trump en dichos estados a las tres de la mañana del 4 de noviembre de 2020, es menor que uno entre mil billones, o uno entre 1.000.000.000.000.000. Para que el exvicepresidente Biden gane esos cuatro estados colectivamente, las probabilidades de que se produzca ese suceso disminuyen a menos de uno entre mil billones elevado a la cuarta potencia». Las impresionantes matemáticas de Paxton asumían que los votos contados en el transcurso del recuento eran estadísticamente independientes, como las tiradas reiteradas de un dado. Sin embargo, los urbanitas tienden a votar de manera diferente que la población suburbana, que a su vez vota de forma distinta a la población rural, y los votantes en persona difieren de quienes emiten su voto por correo (particularmente en 2020, cuando Trump disuadió a sus partidarios de votar por correo). Dentro de cada sector, los votos no son independientes, y las tasas base difieren de

un sector a otro. Dado que los resultados de cada distrito se anuncian a medida que se hallan disponibles, y los votos por correo se recontaron más tarde todavía, conforme se van sumando los diferentes tramos, el continuo recuento que favorece a cada candidato puede subir o bajar, y el resultado final no puede extrapolarse a partir de los resultados provisionales. La necedad se elevó a la cuarta potencia cuando Paxton multiplicó las falsas probabilidades de los cuatro estados, cuyos votos tampoco son independientes: lo que quiera que influya en los votantes del estado de los Grandes Lagos es probable que los influya también en la tierra de los lácteos.[43*]

LA INDEPENDENCIA ESTADÍSTICA está vinculada al concepto de *causalidad*: si un suceso afecta a otro, ambos no son estadísticamente independientes (aunque, como veremos, no viceversa: los sucesos causalmente aislados pueden ser estadísticamente dependientes). Por ese motivo la falacia del jugador es una falacia. Un giro de una ruleta no puede influir en el siguiente, por lo que el jugador que espera que una racha de negros producirá un rojo perderá hasta la camisa: la probabilidad es siempre un poco menor de 0,5 (debido a las casillas verdes con 0 y 00). Esto demuestra que las falacias de la independencia estadística pueden cometerse en ambos sentidos: asumiendo falsamente la independencia (como en la falacia de Meadow) y asumiendo falsamente la dependencia (como en la falacia del jugador).

No siempre resulta evidente si los sucesos son o no independientes. Entre las aplicaciones más célebres de la investigación sobre sesgos cognitivos a la vida cotidiana figura el análisis de Tversky (con el psicólogo social Tom Gilovich) de la «mano caliente» en el baloncesto.[44] Todos los aficionados al básquet saben que, de vez en cuando, un jugador puede estar «en racha», especialmente jugadores como Vinnie *Microondas* Johnson, el escolta de los Detroit Pistons en la década de 1980, que se ganó su apodo porque se calentaba a toda prisa. Contra la incredulidad de todos los aficionados, entrenadores, jugadores y cronistas deportivos, Tversky y Gilovich afirmaban que la mano caliente era una ilusión, una falacia del

jugador a la inversa. Los datos que analizaron sugerían que el resultado de todos los lanzamientos es estadísticamente independiente de la serie de intentos precedentes.

Ahora bien, antes de examinar los datos, no podemos descartar la posibilidad de una mano caliente sobre la base de la plausibilidad causal a la manera en que cabe descartar la falacia del jugador. A diferencia de una ruleta, el cuerpo y el cerebro de un jugador sí que tienen memoria, y dista de ser supersticioso pensar que una eclosión de energía o confianza pueda persistir durante un lapso de unos minutos. Así pues, no se produjo ninguna conmoción en la cosmovisión científica cuando otros estadísticos examinaron por segunda vez los datos y concluyeron que los cerebritos estaban equivocados y que los deportistas estaban en lo cierto: sí que existe una mano caliente en el baloncesto. Los economistas Joshua Miller y Adam Sanjurjo demostraron que, cuando seleccionamos rachas de aciertos o errores en una larga secuencia de datos, el resultado del intento inmediatamente siguiente no es estadísticamente independiente de esa racha. La razón es que, si el intento hubiera resultado exitoso y hubiese continuado la racha, podría haberse contado de entrada como parte de dicha racha. Cualquier intento señalado por haber tenido lugar después de una racha está predispuesto a ser un intento no exitoso: uno que no tenía oportunidad de ser definido como parte de la propia racha. Eso da al traste con los cálculos de lo que cabría esperar del azar, lo que a su vez echa por tierra la conclusión de que los jugadores de baloncesto no son más erráticos que las ruletas.[45]

La *falacia de la falacia* de la mano caliente nos enseña tres lecciones. En primer lugar, los sucesos pueden ser estadísticamente dependientes no solo cuando un suceso afecta causalmente al otro, sino también cuando influye en qué evento es seleccionado para la comparación. En segundo lugar, la falacia del jugador puede ser fruto de una característica no tan irracional de la percepción: cuando buscamos rachas en una larga secuencia de sucesos, una racha de una determinada duración tiene realmente más probabilidades de ser invertida que de continuar. En tercer lugar, la probabilidad puede ser verdadera y profundamente antintuitiva: incluso los expertos en matemáticas pueden meter la pata.

PASEMOS AHORA A LA PROBABILIDAD de una *disyunción* de sucesos, P(A O B). Esta es la probabilidad de A más la probabilidad de B, menos la probabilidad de A Y B. Si los Brown tienen dos hijos, la probabilidad de que al menos uno sea niña —es decir, de que el primero sea niña o el segundo sea niña— es 0,5 + 0,5 – 0,25, es decir, 0,75. Podemos llegar al mismo resultado contando las combinaciones: niño-niña + niña-niña + niña-niño (tres posibilidades) de niño-niña + niño-niño + niña-niño + niña-niña (cuatro oportunidades). O haciendo el recuento de frecuencias: en un gran conjunto de familias con dos hijos, encontraremos que tres cuartas partes tienen al menos una hija.

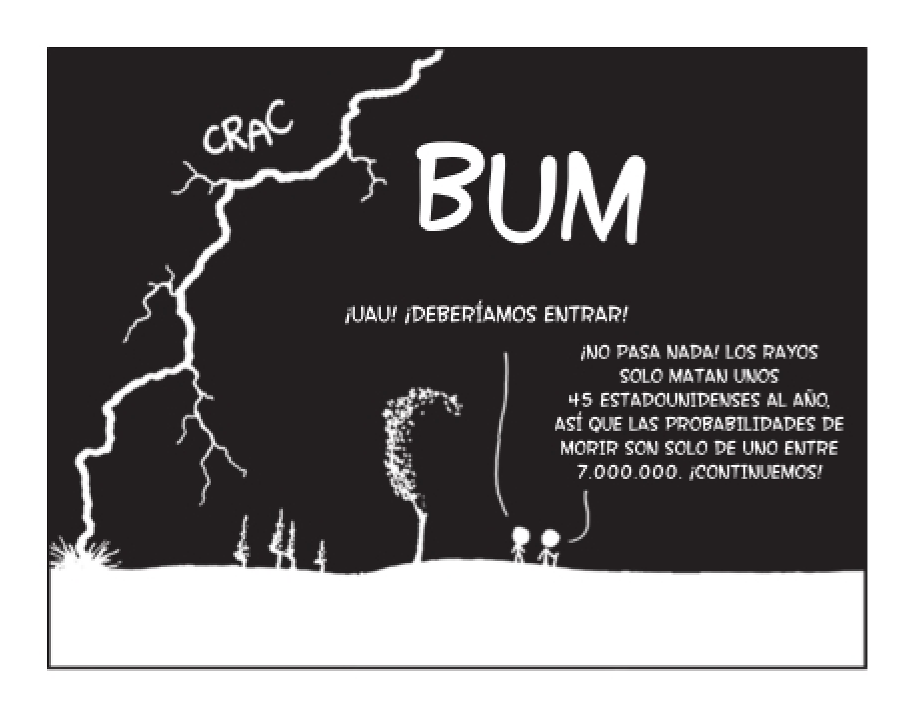
La aritmética de la *o* nos muestra el error del meteorólogo que decía que era seguro que llovería durante el fin de semana porque había un 50 % de probabilidades de que lloviese cada día: al sumar simplemente las dos probabilidades, contabilizó dos veces sin querer los fines de semana en los que llovería *ambos* días, olvidando sustraer 0,25 por la conjunción. Aplicó una regla que funciona para la disyunción exclusiva (XO), a saber: A O B, pero no ambas. Las probabilidades de los sucesos mutuamente excluyentes *pueden* sumarse para obtener la disyunción, y la suma de todos ellos es uno, en efecto. La probabilidad de que un hijo sea un niño (0,5) o una niña (0,5) es su suma, uno, dado que el hijo ha de ser lo uno o lo otro (tratándose de un ejemplo para explicar las matemáticas, he adoptado un binarismo de género, sin considerar los hijos intersexuales). Si olvidamos la diferencia y confundimos los sucesos que se solapan con los mutuamente exclusivos, podemos obtener unos resultados disparatados. Imaginemos que el meteorólogo predijera una probabilidad de lluvia del 0,5 el sábado, el domingo y el lunes, y concluyera que la probabilidad de lluvia a lo largo del largo fin de semana era de 1,5.

La probabilidad del complemento de un suceso, a saber, que no ocurra A, es uno menos la probabilidad de que ocurra. Esto viene bien cuando tenemos que calcular la probabilidad de «al menos un» suceso. ¿Recuerdas a los Brown con su hija, o quizá dos hijas? Dado que tener al menos una hija es lo mismo que no tener dos niños, en lugar de calcular la disyunción (el primer hijo es una niña *o* el segundo hijo es una niña), podríamos haber calculado el complemento de una conjunción: uno, menos la probabilidad

de tener dos niños (que es 0,25), es decir: 0,75. En el caso de dos sucesos, no importa mucho qué fórmula usemos. Ahora bien, cuando tenemos que calcular la probabilidad de al menos una A en un conjunto grande, la regla de la disyunción requiere el tedio de sumar y restar muchas combinaciones. Resulta mucho más fácil calcularla como la probabilidad de «no todos NO A», que es simplemente uno menos un gran producto.

Supongamos, por ejemplo, que cada año hay un 10 % de probabilidades de que estalle una guerra. ¿Cuáles son las probabilidades de que estalle al menos una guerra a lo largo de una década? (Supongamos que las guerras sean independientes y no contagiosas, lo cual parece ser cierto).[46] En lugar de sumar la probabilidad de que estalle una guerra en el año 1 más la probabilidad de que estalle en el año 2 menos la probabilidad de que estalle tanto en el año 1 como en el año 2, y así sucesivamente para todas las combinaciones, podemos calcular simplemente la probabilidad de que no estalle *ninguna* guerra durante *todos* los años y restársela a uno. Esa es simplemente la probabilidad de que no estalle la guerra en un año dado, 0,9, multiplicado por sí mismo para cada uno de los años restantes (0,9 × 0,9 × ... 0,9; o $0{,}9^{10}$, que es igual a 0,35), que, cuando se lo restamos a uno, resulta 0,65.

Y LLEGAMOS POR FIN a una probabilidad *condicional*: la probabilidad de A dado B, expresada como P(A | B). Una probabilidad condicional es conceptualmente simple: es sencillamente la probabilidad del *entonces* en un *si... entonces*. También es simple en términos aritméticos: es simplemente la probabilidad de A Y B dividida entre la probabilidad de B. No obstante, es la fuente de interminables confusiones, errores garrafales y paradojas en los razonamientos relativos a la probabilidad, empezando por el desventurado personaje de la viñeta de *XKCD* de la siguiente página.[47] Su error consiste en confundir la probabilidad simple o *tasa base* de muertes a causa de los rayos, P(alcanzado por rayo), con la probabilidad *condicional* de una muerte por un rayo, dado que uno está fuera durante una tormenta eléctrica, P(alcanzado por rayo | fuera en tormenta).



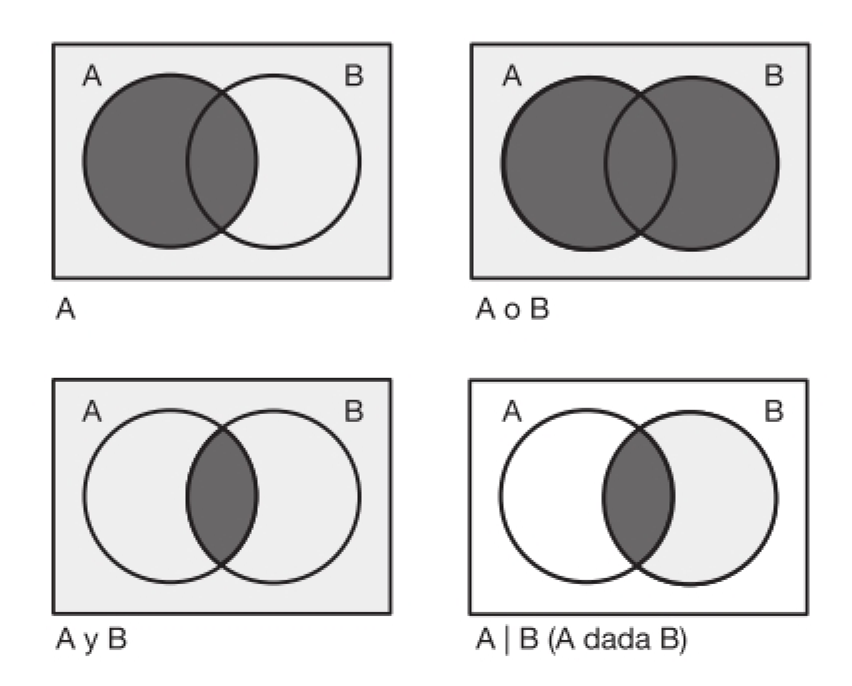
LA TASA ANUAL DE MORTALIDAD ENTRE LAS PERSONAS QUE CONOCEN ESA ESTADÍSTICA ES UNA DE CADA SEIS.

xkcd.com

Aunque la aritmética de una probabilidad condicional es simple, resulta antintuitiva hasta que la concretamos y la tornamos visualizable (como siempre). Observemos los diagramas de Venn de la página siguiente, en los que el tamaño de una región de la página corresponde al número de resultados. El rectángulo, con un área de uno, abarca todas las posibilidades. El círculo A comprende todas las A, y la figura superior izquierda muestra que la probabilidad de A corresponde a su área (oscura) como una proporción del rectángulo entero (claro): otra forma de expresar el número de casos dividido entre el número de oportunidades. La figura superior derecha muestra la probabilidad de A O B, que es el área oscura total, a saber: el área de A más el área de B, sin el doble cómputo de la porción que comparten en el medio, es decir, la probabilidad de A Y B. Esa porción, P(A Y B), se muestra en el diagrama inferior izquierdo.

El diagrama inferior derecho explica lo que sucede con las probabilidades condicionales. Indica que deberíamos ignorar el vasto espacio de todo cuanto puede ocurrir, representado en color blanco, y centrar nuestra atención exclusivamente en los casos en los que sucede B, el círculo sombreado. Examinemos ahora en cuántos de *esos* casos también

ocurre A: el tamaño de la porción A Y B como una proporción del tamaño del círculo B. De todos los interludios en los que la gente camina en medio de una tormenta eléctrica (B), ¿qué proporción de ellos desembocan en la caída de un rayo (A Y B)? Por eso calculamos el condicional, P(A | B), dividiendo la conjunción, P(A Y B) entre la tasa base, P(B).



Veamos un ejemplo. Los Gray tienen dos hijos. El mayor es una niña. Sabiendo esto, ¿cuál es la probabilidad de que ambos sean niñas? Traduzcamos la pregunta a una probabilidad condicional, a saber: la probabilidad de que el primero sea niña y el segundo sea niña dado que el primero es niña, o, en una notación sofisticada, P(1.º = niña Y 2.º = niña | 1.º = niña). La fórmula nos dice que dividamos la conjunción, que ya hemos calculado que es 0,25, entre la probabilidad simple para el segundo hijo, 0,5, y obtenemos 0,5. O, pensando en términos clásicos y concretos: niña-niña (una posibilidad) dividida entre niña-niña y niña-niño (dos oportunidades), igual a un medio.

Las probabilidades condicionales añaden algo de precisión al concepto de *independencia estadística*, que dejé en suspenso en la subsección precedente. Podemos definir ahora el concepto: A y B son independientes si, para toda B, la probabilidad de A dada B es la misma que la probabilidad

total de A (y lo mismo para B). Pues bien, ¿recuerdas la multiplicación ilícita de las probabilidades para la conjunción de sucesos cuando estos no son independientes? ¿Qué se supone que hemos de hacer en lugar de ello? Es fácil: la probabilidad de la conjunción de A Y B cuando *no* son independientes es la probabilidad de A veces la probabilidad de B dada A, esto es, P(A) × P(B | A).

¿Por qué estoy insistiendo en el concepto de *probabilidad condicional* con todas estas representaciones sinónimas: la prosa inglesa o castellana, su equivalente lógico, la fórmula matemática, los diagramas de Venn, el recuento de las posibilidades? El motivo es que la probabilidad condicional es semejante fuente de confusión que las explicaciones nunca son excesivas.[48]

Si no me crees, consideremos a los White, otra familia con dos hijos. Al menos uno de ellos es una niña. ¿Cuál es la probabilidad de que ambos sean niñas, esto es, la probabilidad condicional de dos niñas dada al menos una niña, o P(1.º = niña Y 2.º = niña | 1.º = niña O 2.º = niña)? Son tan pocos los que llegan a la respuesta correcta que los estadísticos la denominan *paradoja del niño o niña*. La gente tiende a decir 0,5; la respuesta correcta es 0,33. En este caso, el pensamiento concreto puede conducir a la respuesta incorrecta: la gente visualiza a una niña mayor, advierte que esta podría tener o bien una hermana menor o bien un hermano menor, y piensa que la hermana es una posibilidad entre estas dos. Se olvida así que hay otra forma de tener al menos una niña: esta podría ser la hermana menor. Enumerando adecuadamente las posibilidades, tenemos niña-niña (una) dividido entre [niña-niña + niña-niño + niño-niña] (tres), que es igual a un tercio. O, empleando la fórmula, dividimos 0,25 (niña Y niña) entre 0,75 (niña O niña).

La paradoja del niño o niña no es un simple truco basado en la redacción. Surge de una falta de imaginación al enumerar las posibilidades, y aparece bajo muchos disfraces, incluido el dilema de Monty Hall. He aquí un equivalente más simple pero exacto.[49] Algunos tahúres callejeros se ganan la vida enganchando a los transeúntes para que jueguen a las tres cartas en un sombrero. El fullero les muestra una carta que es roja por ambas caras, otra que es blanca por ambas caras y una tercera que es roja

por un lado y blanca por el otro. Las mezcla en el sombrero, saca una, observa que la cara es (pongamos por caso) roja y apuesta al transeúnte a que la otra cara es también roja (este le paga un dólar si es roja, y él paga un dólar al transeúnte si es blanca). Se trata de una apuesta de bobos: las probabilidades de que sea roja son dos de tres. Los palurdos cuentan mentalmente las cartas en lugar de las *caras* de cartas, olvidando que son dos las formas en la que la carta es toda roja, de haber sido elegida, podría haber aparecido con la cara roja hacia arriba.

¿Y recuerdas al tipo que embarcó en el avión con su propia bomba? Calculó la probabilidad total de que un avión llevase a bordo dos bombas. Ahora bien, al llevar su propia bomba, ya había descartado la mayoría de las posibilidades en el denominador. El número del que debería preocuparse es la probabilidad condicional de que un avión tenga dos bombas *dado que* ya tiene una bomba, la suya (que tiene una probabilidad de uno). Ese condicional es la probabilidad de que alguien más tenga una bomba multiplicada por uno (la conjunción de su bomba y la del otro tipo), dividida entre uno (su bomba), que resulta, por supuesto, la probabilidad de que otra persona tenga una bomba, regresando al punto de partida. El chiste se utilizó con buenos resultados en *The World According to Garp (El mundo según Garp).* Los Garp están viendo una casa cuando un pequeño avión se estrella contra ella. Garp dice: «Nos quedamos con la casa. La probabilidad de que se estrelle contra esta casa otro avión es astronómica».[50]

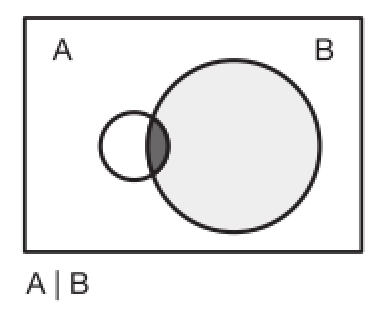
Olvidarse de condicionar una probabilidad de la tasa base en función de las especiales circunstancias que concurren (la tormenta eléctrica, la bomba que subes a bordo) es una metedura de pata probabilística habitual. Durante el juicio de 1995 de O. J. Simpson, la estrella del fútbol acusada de asesinar a su esposa Nicole, un fiscal llamó la atención sobre su historial de palizas. Un miembro del Dream Team de abogados defensores de Simpson respondió que muy pocos maltratadores llegan a matar a sus mujeres, quizá uno de cada dos mil quinientos. Una profesora inglesa, Elaine Scarry, detectó la falacia. Nicole Simpson no era simplemente una antigua víctima de agresiones. Era una víctima de agresiones *que había sido degollada*. La

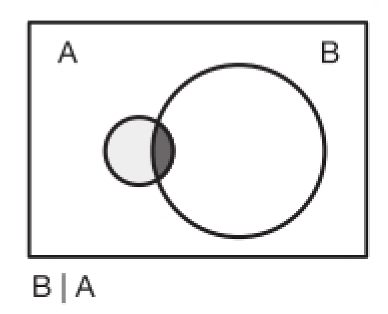
estadística relevante es la probabilidad condicional de que alguien matara a su mujer dado que él había maltratado a su mujer *y que su mujer había sido asesinada por alguien*. Esa probabilidad es de ocho de nueve.[51]

EL OTRO ERROR COMÚN con la probabilidad condicional consiste en confundir la probabilidad de A dada B con la probabilidad de B dada A, el equivalente estadístico de afirmar el consecuente (pasar de «si P entonces Q» a «si Q entonces P»).[52] ¿Recuerdas a Irwin, el hipocondríaco que sabía que tenía una enfermedad hepática porque sus síntomas encajaban perfectamente con la lista, a saber, no tener ningún síntoma? Irwin confundía la probabilidad de la ausencia de síntomas dada la enfermedad hepática, que es alta, con la probabilidad de la enfermedad hepática dada la ausencia de síntomas, que es baja. Ello es debido a que la probabilidad de la enfermedad hepática (su tasa base) es baja, y la probabilidad de la ausencia de malestar es elevada.

Las probabilidades condicionales no pueden invertirse cada vez que difieren las tasas base. Tomemos un ejemplo de la vida real, el hallazgo de que un tercio de los accidentes mortales ocurren en el hogar, lo cual inspiró el titular «Las viviendas particulares son sitios peligrosos». El problema estriba en que la casa es el lugar donde pasamos la mayor parte de nuestro tiempo, por lo que, incluso si los hogares no son particularmente peligrosos, muchos de nuestros accidentes se producen en ellos porque en ellos nos sucede mucho de *todo*. El autor del titular confundía la probabilidad de que estuviéramos en casa dado que había ocurrido un accidente mortal (la estadística referida), con la probabilidad de que ocurra un accidente mortal dado que estamos en casa, que es la tendencia que interesa a los lectores. Podemos comprender el problema de un modo más intuitivo observando el siguiente diagrama, donde las tasas base se reflejan en los tamaños relativos de los círculos (por ejemplo, siendo A los días con accidentes mortales y B los días en casa). El diagrama izquierdo muestra la probabilidad de A dada B (la probabilidad de un accidente mortal dado que uno está en casa); es el área del sector oscuro (A Y B) como una proporción del gran círculo claro (B, estar en casa), que es pequeña. El diagrama de la derecha muestra la probabilidad de B dada A (de estar en casa dado que se ha producido un

accidente mortal); es el área de ese mismo sector oscuro, pero esta vez como una proporción del pequeño círculo claro, accidentes mortales, y es mucho mayor.





Una de las razones por las que resulta tan fácil entender al revés las probabilidades condicionales es que el inglés y el castellano son ambiguos respecto de lo que se pretende. «La probabilidad de que un accidente tenga lugar en casa es 0,33» podría significar «una proporción de los accidentes» o «una proporción del tiempo pasado en casa». La diferencia puede perderse en la traducción y generar falsas estimaciones de tendencias. En una mayoría de los accidentes de bicicleta hay niños involucrados, por lo que obtenemos el titular «Los niños corren más riesgo en bicicleta», que da a entender que los niños son más imprudentes, mientras que de hecho puede que sean simplemente ciclistas más ávidos. Y, en lo que los estadísticos llaman la falacia del fiscal, el fiscal del distrito anuncia que la probabilidad de que el grupo sanguíneo de la víctima coincida por casualidad con el de la ropa del acusado es solo del 3 %, y concluye que la probabilidad de que el acusado sea culpable es del 97 %. Ha confundido (y confía en que los miembros del jurado confundan) la probabilidad de una coincidencia dado que el acusado es inocente con la probabilidad de que el acusado sea inocente dada una coincidencia.[53] Cómo efectuar adecuadamente los cálculos aritméticos es el tema del próximo capítulo, el razonamiento bayesiano.

Las ambigüedades en la probabilidad condicional pueden ser incendiarias. En 2019, un par de científicos sociales causaron furor cuando publicaron un estudio en la prestigiosa revista *Proceedings of the National Academy of Sciences* que, citando números como los mencionados en una

sección anterior, afirmaba que la policía tenía más probabilidades de disparar a los blancos que a los negros, en contra de la suposición habitual del sesgo racial. Los críticos señalaron que esa conclusión pertenecía a la probabilidad de que alguien sea negro dado que ha recibido disparos, que es en efecto más baja que la probabilidad correspondiente para los blancos, pero solo porque en el país hay menos negros que blancos, una diferencia en las tasas base. Si la policía *tuviera* prejuicios raciales, la tendencia se manifestaría como una probabilidad más elevada de que alguien reciba disparos dado que es negro, y los datos sugieren que la probabilidad es más alta en efecto. Aunque los autores originales advirtieron que la tasa base apropiada no resulta evidente —¿debería ser la proporción de negros en la población o en los encuentros con la policía?—, se percataron de que habían formado un lío tan grande en su manera de formular las probabilidades que retiraron formalmente el artículo.[54]

¿Y el papa procedente del espacio exterior? Eso es lo que obtenemos cuando confundimos la probabilidad de ser papa, dado que alguien es humano, con la probabilidad de que alguien sea humano, dado que es el papa.[55]

PROBABILIDADES PREVIAS Y *POST HOC*

Un hombre se prueba un traje y le dice al sastre: «Necesito que acorte esta manga». El sastre contesta: «No, doble el codo así. Ya ve cómo sube la manga». El cliente replica: «Sí, de acuerdo, pero cuando doblo el codo, se sube el cuello por la nuca». El sastre dice: «Pues levante la cabeza y échela hacia atrás. Perfecto». El hombre exclama: «¡Pero ahora el hombro izquierdo está siete centímetros más bajo que el derecho!». El sastre comenta: «Ningún problema. Doble la cintura y se iguala». El hombre sale de la tienda llevando el traje, con el codo derecho sobresaliente, la cabeza inclinada hacia atrás y el torso doblado hacia la izquierda, caminando con paso espasmódico. Un par de peatones se cruzan con él. El primero dice: «¿Has visto a ese pobre discapacitado? Me da mucha lástima». El segundo responde: «Sí, pero su sastre es un genio, ¡el traje le sienta como anillo al dedo!».

El chiste ilustra otra familia de meteduras de pata probabilísticas: la confusión de los juicios previos con los *post hoc* (también llamados *a priori* y *a posteriori*). La confusión se conoce a veces como la falacia del francotirador de Texas, por el tirador que dispara una bala a un lado del granero y luego pinta una diana alrededor del agujero. En el caso de la probabilidad, hay una gran diferencia si el denominador de la fracción (el número de oportunidades de que se produzca un suceso) se cuenta o no independientemente del numerador, los sucesos de interés. El sesgo de confirmación, comentado en el capítulo 1, provoca el error: una vez que esperamos un patrón, buscamos ejemplos e ignoramos los contraejemplos. Si tomamos nota de las predicciones de un vidente que los acontecimientos confirman, pero no dividimos entre el número total de predicciones, correctas e incorrectas, podemos obtener cualquier probabilidad que deseemos. Como observó Francis Bacon en 1620, tal es la forma de todas las supersticiones, ya sea en la astrología, ya en los sueños, los augurios o los juicios divinos.

O en los mercados financieros. Un asesor de inversiones sin escrúpulos con una lista de correo de cien mil personas envía a la mitad de la lista un boletín informativo que predice la subida del mercado, y una versión a la otra mitad que predice su caída. Al final de cada trimestre descarta los nombres que recibieron la predicción equivocada y repite el proceso. Al cabo de dos años, consigue la suscripción de 1.562 destinatarios que están asombrados por su historial de predicciones sobre el mercado durante ocho trimestres consecutivos.[56]

Aunque este timo es ilegal si se lleva a cabo a sabiendas, cuando se practica ingenuamente es la savia de la industria financiera. Los comerciantes son rápidos como un rayo a la hora de agarrar las gangas, por lo que muy pocos seleccionadores de acciones pueden superar los resultados de una cesta de valores mecánica. Una excepción fue Bill Miller, ungido por CNNMoney.com en 2006 como «el mejor administrador de dinero de nuestro tiempo» por batir al índice del mercado bursátil S&P 500 durante quince años consecutivos. ¿Cuán impresionante es esto? Cabría pensar que si un administrador tiene las mismas probabilidades de obtener mejores o peores resultados que el índice en cualquier año, las posibilidades

de que eso suceda por azar son solo de una entre 32.768 (2^{15}). Pero Miller fue señalado *después* de desplegar su asombrosa racha. Como señaló el físico Leonard Mlodinow en *The Drunkard's Walk: How Randomness Rules Our Lives (El andar del borracho: cómo el azar gobierna nuestras vidas)*, el país cuenta con más de seis mil gestores de fondos, y los fondos de inversión mobiliaria modernos existen desde hace cuarenta años aproximadamente. No es improbable en absoluto que *algún* gestor tuviera *alguna vez* una racha ganadora de quince años a lo largo de esos cuarenta años; es una probabilidad de tres sobre cuatro. El titular de *CNNMoney* podría haber rezado «La esperada racha de quince años sucede finalmente: Bill Miller es el afortunado». Como era de esperar, la suerte de Miller se agotó y, en los dos años siguientes, el mercado «lo pulverizó con comodidad».[57]

Además del sesgo de confirmación, un factor esencial de las falacias de la probabilidad *post hoc* es nuestra incapacidad de apreciar cuántas oportunidades existen de que se produzcan las coincidencias. Cuando se nos permite identificarlas *post hoc*, las coincidencias no son improbables, sino altamente probables. En una de sus columnas del *Scientific American*, el matemático recreativo Martin Gardner preguntaba: «¿Te darías cuenta si la placa de matrícula de un coche justo delante de ti constara de dígitos que, leídos hacia atrás, dieran tu número de teléfono? ¿Quién excepto un numerólogo o un logófilo vería las letras U, S, A simétricamente situadas en *Louisiana* o al final de *John Philip Sousa*, el nombre del compositor de una de las más grandes marchas patrióticas? Hace falta una mente muy particular para descubrir que Newton nació el mismo año en que murió Galileo, o que Bobby Fischer (Pescador) nació bajo el signo de Piscis (el pez)».[58] Pero esos numerólogos y mentes tan particulares existen, y con su puntería *post hoc* pueden tejerse presuntuosas teorías. El psicoanalista Carl Jung propuso una fuerza mística llamada sincronicidad para explicar la quintaesencia de lo que no necesita explicación: la prevalencia de la coincidencia en el mundo.

Cuando yo era niño, lo que hoy llamamos memes circulaban en tebeos y revistas populares. Uno muy difundido fue una lista de las increíbles similitudes entre Abraham Lincoln y John F. Kennedy. Tanto el honesto

Abe como JFK fueron elegidos para el Congreso en el año 46 y para la presidencia en el 60. Ambos recibieron disparos en la cabeza un viernes en presencia de sus esposas respectivas. Lincoln tenía una secretaria llamada Kennedy; Kennedy tenía una secretaria llamada Lincoln. Ambos fueron sucedidos por un Johnson nacido en el año 8. Los asesinos de ambos nacieron en el 39 y tenían tres nombres que sumaban un total de quince letras. John Wilkers Booth huyó desde un teatro y fue atrapado en un almacén; Lee Harvey Oswald corrió desde un almacén y fue atrapado en un teatro. ¿Qué nos dicen estos espeluznantes paralelismos? Con todo el respeto debido al doctor Jung, absolutamente nada, aparte de que las coincidencias se producen con más frecuencia de lo que aprecian nuestras mentes poco instruidas en estadística. Por no mencionar el hecho de que, cuando se advierten coincidencias escalofriantes, estas tienden a adornarse (Lincoln no tenía una secretaria llamada Kennedy), al tiempo que se ignoran las incómodas «no coincidencias» (como sus diferentes días, meses y años de nacimiento y muerte).

Los científicos no son inmunes a la falacia del francotirador de Texas. Esta es una de las explicaciones para la crisis de replicabilidad que ha sacudido la epidemiología, la psicología social, la genética humana y otros campos en la década de 2010.[59] Piensa en todos los alimentos que son buenos para ti que solían ser malos para ti, en el medicamento milagroso que resulta no funcionar mejor que el placebo, el gen de este o aquel rasgo que en realidad era ruido en el ADN, los ingeniosos estudios que demostraban que los individuos dejan mejores propinas cuando se cuelgan dos ojos dibujados en la pared de una cafetería, y que caminan más despacio hacia el ascensor después de realizar un experimento en el que se les mostraban palabras asociadas con la vejez.

No es que los investigadores falsificaran sus datos. Es que llevaban a cabo prácticas investigadoras que hoy sabemos que son cuestionables, el jardín de los senderos que se bifurcan y el *p-hacking* o dragado de datos (que se refiere al umbral de probabilidad, *p*, que se considera «estadísticamente significativo»).[60] Imaginemos a un científico que lleva a cabo un laborioso experimento y obtiene datos que son lo contrario del «eureka». Antes de abandonar el experimento, puede sentir la tentación de

preguntarse si el efecto se produce realmente, pero solo en el caso de los hombres, o solo con las mujeres, o si se rechazan los datos anómalos de los participantes que se habían distraído, o si se excluyen los años locos de Trump, o si se cambia a una prueba estadística que se fije en la clasificación de los datos más que en sus valores hasta el último decimal. O bien podemos continuar sometiendo al test a los participantes hasta que aparezca el precioso asterisco en el impreso estadístico, asegurándonos de dejarlo mientras llevamos la delantera.

Ninguna de estas prácticas es intrínsecamente irrazonable si puede justificarse antes de la recogida de datos. Ahora bien, si estos son probados después de los hechos, es probable que alguna combinación saque provecho del azar y arroje un resultado espurio. La trampa es inherente a la naturaleza de la probabilidad y se conoce desde hace décadas; recuerdo que, cuando estudié Estadística en 1974, fui advertido del peligro del *fisgoneo de datos*. No obstante, hasta fechas recientes, eran pocos los científicos que captaban intuitivamente que una pizca de fisgoneo de datos podía abocar a un cargamento de errores. Mi profesor sugería medio en broma que se exigiera a los científicos que escribiesen sus hipótesis y sus métodos en un papel antes de realizar un experimento y lo custodiasen en una caja fuerte, que abrirían y mostrarían a los evaluadores una vez concluido el estudio.[61] El único problema, advertía, estribaba en que los científicos podían mantener en secreto varias cajas fuertes y luego abrir aquella que sabían que «predecía» los datos. Con el advenimiento de la web, el problema se ha solucionado, y la vanguardia de la metodología científica consiste en «prerregistrar» los detalles de un estudio en un registro público, que los evaluadores y editores pueden comprobar, a fin de detectar posibles tejemanejes *post hoc*.[62]

EXISTE UN TIPO DE ILUSIÓN de la probabilidad *post hoc* tan común que tiene su propio nombre: la *ilusión de agrupación*.[63] Se nos da bien percibir conjuntos de cosas o sucesos muy compactos, porque a menudo forman parte de un único acontecimiento: un perro ladrador que no se calla, un sistema climático que empapa una ciudad durante varios días, un ladrón en acción que roba varias tiendas de una misma manzana. Pero no todos los

agrupamientos tienen una causa principal; de hecho, la mayoría de ellos no la poseen. Cuando hay muchos sucesos, resulta inevitable que los unos deambulen por los barrios de los otros y se codeen con ellos, a menos que algún proceso no aleatorio trate de mantenerlos separados.

La ilusión de agrupación nos hace pensar que los procesos aleatorios son no aleatorios y viceversa. Cuando Tversky y Kahneman mostraron a la gente (estadísticos incluidos) los resultados de series reales de lanzamientos de monedas, como CrCrCaCaCrCaCrCrCrCr, que inevitablemente tienen rachas de caras o cruces consecutivas, los individuos pensaban que la moneda estaba trucada. Solo decían que una moneda parecía equilibrada si *estaba* trucada de hecho para evitar las rachas, como CaCrCaCrCrCaCrCaCaCr, que «parece» aleatoria aun cuando no lo es.[64] Yo fui testigo de una ilusión similar cuando trabajaba en un laboratorio de percepción auditiva. Los participantes tenían que detectar sonidos tenues, que se presentaban en momentos aleatorios a fin de que no pudieran adivinar cuándo sonaría un tono. Algunos decían que el generador de sucesos aleatorios debía de estar averiado porque los tonos llegaban en ráfagas. No se daban cuenta de que así es como suena la aleatoriedad.

Los agrupamientos fantasmas surgen asimismo en el espacio. Las estrellas que forman Aries, Leo, Cáncer, Virgo, Sagitario y otras constelaciones no son vecinas en ninguna galaxia, sino que se encuentran diseminadas por el cielo nocturno desde nuestro punto de vista terrestre y solo son agrupadas en formas por nuestros cerebros buscadores de patrones. También aparecen agrupaciones espurias en el calendario. La gente se sorprende al enterarse de que, si hay veintitrés personas en una habitación, las probabilidades de que dos de ellas compartan el día de nacimiento superan el 50 %. Con cincuenta y siete personas en la habitación, las probabilidades ascienden hasta el 99 %. Aunque es improbable que haya alguien en la habitación que comparta *mi* día de nacimiento, no estamos buscando coincidencias conmigo, ni con nadie seleccionado *a priori*. Estamos contando coincidencias *post hoc*, y la coincidencia puede producirse de 366 maneras.

La ilusión de agrupación, al igual que otras falacias *post hoc* en probabilidad, es la fuente de numerosas supersticiones: que las cosas malas vienen de tres en tres, que hay gente que nace con mala estrella o que un *annus horribilis* significa que el mundo se está desmoronando. Cuando una serie de plagas se abate sobre nosotros, eso no quiere decir que haya un Dios que nos esté castigando por nuestros pecados o poniendo a prueba nuestra fe. Significa que no hay un Dios que las esté espaciando.

INCLUSO PARA AQUELLOS que entienden las matemáticas del azar en toda su fastidiosa antintuitividad, una racha de buena suerte puede apoderarse de la imaginación. Las probabilidades subyacentes determinarán cuánto tiempo se espera que dure por término medio una racha, pero el momento exacto en el que se agota la suerte es un misterio insondable. Esta tensión fue explorada en mi ensayo favorito del paleontólogo, escritor científico y aficionado al béisbol Stephen Jay Gould.[65]

Gould comentaba uno de los mayores logros deportivos, la racha de bateo de cincuenta y seis partidos de Joe DiMaggio en 1941. Explicaba que la racha era estadísticamente extraordinaria incluso dado el elevado promedio de bateo de DiMaggio y el número de oportunidades de que se hubieran producido rachas en la historia del deporte. El hecho de que DiMaggio se beneficiara de algunas rachas de suerte a lo largo del camino no disminuye la hazaña, sino que la ejemplifica, porque ninguna racha larga, por muy impulsada que esté por las probabilidades favorables, puede desplegarse jamás sin ella. Gould explica nuestra fascinación por las rachas de suerte:

> Las estadísticas de buenas rachas y crisis, adecuadamente entendidas, nos enseñan una lección importante acerca de la epistemología y la vida en general. La historia de una especie, o cualquier fenómeno natural que requiera una continuidad ininterrumpida en un mundo de problemas, funciona como una racha de bateo. Todos son partidos de un jugador que juega con una apuesta limitada contra una casa con infinitos recursos. El jugador debe acabar arruinándose. Su meta solamente puede ser permanecer todo el tiempo posible, divertirse durante el trayecto y, si resulta ser asimismo un agente moral, preocuparse de mantener el rumbo con honor [...].
>
> La racha de bateo de DiMaggio es la más excelente de las leyendas legítimas porque encarna la esencia de la batalla que define verdaderamente nuestra vida. DiMaggio activó el sueño más grande y más inalcanzable de toda la humanidad, la esperanza y la quimera de todos los sabios y chamanes: engañó a la muerte, al menos durante algún tiempo.

CAPÍTULO

5

Creencias y evidencias
(Razonamiento bayesiano)

Las afirmaciones extraordinarias requieren evidencias extraordinarias.

CARL SAGAN

Una excepción alentadora al desdén por la razón en tantos de nuestros discursos *online* es el surgimiento de una Comunidad de la Racionalidad cuyos miembros se afanan por estar «menos equivocados», compensando sus sesgos cognitivos y abrazando los estándares del pensamiento crítico y la humildad epistémica.[1] La introducción a uno de sus tutoriales en línea puede servir de introducción al tema de este capítulo:[2]

> La regla de Bayes o el teorema de Bayes es la ley de la probabilidad que gobierna la *fuerza de las evidencias*; la regla establece *cuánto* hemos de revisar nuestras probabilidades (cambiar de parecer) cuando conocemos un nuevo hecho u observamos una nueva evidencia.
>
> Puede interesarte conocer la regla de Bayes si eres:
>
> • Un profesional que maneja estadísticas, como un científico o un médico.
> • Un programador informático que trabaja en el aprendizaje automático.
> • Un ser humano.

Sí, un ser humano. Muchos racionalistas creen que la regla de Bayes es uno de los modelos normativos más frecuentemente incumplidos en los razonamientos cotidianos y que, si se apreciara mejor, podría suponer un fuerte impulso para la racionalidad pública. En las últimas décadas, el pensamiento bayesiano ha disparado su prominencia en todos los ámbitos científicos. Aunque pocos legos pueden nombrarlo o explicarlo, han sentido su influencia en el término de moda *priors* («probabilidades previas» o *a priori*), que hace referencia a una de las variables del teorema.

Un caso paradigmático del razonamiento bayesiano es el diagnóstico médico. Supongamos que la prevalencia del cáncer de mama en la población de mujeres es del 1 %. Supongamos que la sensibilidad de una prueba de cáncer de mama (su tasa de positivos verdaderos) es del 90 %. Supongamos que su tasa de falsos positivos es del 9 %. Una mujer da positivo en el test. ¿Cuál es la probabilidad de que tenga la enfermedad?

La respuesta más popular de una muestra de médicos dadas estas cifras oscilaban entre el 80 y el 90 %.[3] La regla de Bayes nos permite calcular la respuesta correcta: 9 %. Así es: los profesionales a quienes confiamos nuestras vidas meten la pata en la tarea básica de interpretar una prueba médica, y no levemente. Piensan que existe casi un 90 % de probabilidades

de que esa mujer tenga cáncer, cuando en realidad hay un 90 % de probabilidades de que no lo tenga. Imagínate tu reacción emocional al oír una u otra cifra, y considera cómo sopesarías tus opciones en respuesta a ella. Por este motivo, como ser humano, te interesa conocer el teorema de Bayes.

La toma de decisiones arriesgadas requiere tanto valorar las posibilidades («¿tengo cáncer?») como sopesar las consecuencias de cada opción («si no hago nada y tengo cáncer, podría morir; si me someto a cirugía y no tengo cáncer, sufriré un dolor y una desfiguración innecesarios»). En los capítulos 6 y 7 exploraremos cuál es la mejor manera de tomar decisiones importantes cuando conocemos las probabilidades, pero el punto de partida han de ser las probabilidades mismas: dadas las evidencias, ¿cuál es la probabilidad de que un estado de cosas sea verdadero?

Por mucho que nos asuste la palabra *teorema*, la regla de Bayes es bastante simple y, como veremos al final del capítulo, puede tornarse intuitiva. La gran idea del reverendo Thomas Bayes (1701-1761) consistía en que el grado de creencia en una hipótesis puede cuantificarse como una probabilidad (este es el sentido subjetivista del término *probabilidad* que conocimos en el capítulo anterior). Llamémoslo P(hipótesis), la probabilidad de una hipótesis, esto es, nuestro grado de creencia en que esta es verdadera (en el caso del diagnóstico médico, la hipótesis es que el paciente tiene la enfermedad). Obviamente, nuestra creencia en cualquier idea debería depender de las evidencias. En la jerga probabilística, podemos decir que nuestra creencia debería estar *condicionada* a las evidencias. Lo que buscamos es la probabilidad de una hipótesis dados los datos, o P(hipótesis | datos). Esta se denomina *probabilidad posterior* o *a posteriori*, es decir, nuestra creencia en una idea tras haber examinado las evidencias.

Si has dado este paso conceptual, ya estás preparado para la regla de Bayes, porque esta es solamente la fórmula de la probabilidad condicional o condicionada, que conocimos en el capítulo anterior, aplicada a la creencia y las evidencias. Recordemos que la probabilidad de A dada B es la probabilidad de A Y B dividida entre la probabilidad de B. Por tanto, la probabilidad de una hipótesis dados los datos (lo que estamos buscando) es

la probabilidad de la hipótesis *y* los datos (digamos que el paciente tiene la enfermedad *y* el resultado de la prueba es positivo), dividida entre la probabilidad de los datos (la proporción total de pacientes que dan positivo, tanto sanos como enfermos). Expresado como una ecuación: P(hipótesis | datos) = P(hipótesis Y datos) / P(datos). Un recordatorio más del capítulo 4: la probabilidad de A Y B es la probabilidad de A veces la probabilidad de B dada A. Hagamos esa sencilla sustitución y obtendremos la regla de Bayes:

$$P(\text{hipótesis} \mid \text{datos}) = \frac{P(\text{hipótesis}) \times P(\text{datos} \mid \text{hipótesis})}{P(\text{datos})}$$

¿Qué significa esto? Recordemos que P(hipótesis | datos) —la expresión del lado izquierdo— es la probabilidad posterior o *a posteriori*: nuestra creencia actualizada en la hipótesis después de haber examinado las evidencias. Esta podría ser nuestra confianza en un diagnóstico de enfermedad tras haber visto los resultados de las pruebas.

P(hipótesis) —en el lado derecho— significa la probabilidad *previa* o *a priori*, nuestra creencia en la hipótesis *antes* de examinar los datos: cuán plausible o bien establecida era, qué estaríamos obligados a conjeturar si no tuviéramos conocimiento alguno de los datos disponibles. En el caso de una enfermedad, esta podría ser su prevalencia en la población, la tasa base.

P(datos | hipótesis) se denomina *verosimilitud*. En el mundo de Bayes, *verosimilitud (likelihood)* no es sinónimo de *probabilidad*, sino que se refiere a cuán probables resultarían los datos *si* la hipótesis fuese verdadera.[4] Si alguien tiene en efecto la enfermedad, ¿cuán probable es que muestre un síntoma dado o que dé positivo en una prueba?

Y P(datos) es la probabilidad de los datos en general, tanto si la hipótesis es verdadera como si es falsa. A veces se designa como *probabilidad marginal*, no en el sentido de «menor», sino en el sentido de sumar los totales de cada fila (o de cada columna) a lo largo del margen de la tabla: la probabilidad de obtener esos datos cuando la hipótesis es verdadera más la probabilidad de obtener esos datos cuando la hipótesis es

falsa. En el caso del diagnóstico médico, se refiere a la proporción de *todos* los pacientes que tienen un síntoma, u obtienen un resultado positivo, tanto sanos como enfermos.

Sustituyendo el álgebra por la mnemotecnia, la regla de Bayes se convierte en:

$$\text{Probabilidad } a\ posteriori = \frac{\text{probabilidad } a\ priori \times \text{verosimilitud de los datos}}{\text{Habitualidad de los datos}}$$

Traducido al castellano, resulta que «nuestra creencia en una hipótesis después de examinar las evidencias debería ser nuestra creencia previa en la hipótesis, multiplicada por cuán probables serían las evidencias *si* la hipótesis fuese verdadera, escalada por lo comunes que son esas evidencias en general».

Y traducido al sentido común, funciona de este modo. Ahora que he visto las evidencias, ¿cuánto debería creer la idea? En primer lugar, créela más si la idea estaba bien respaldada, era creíble o plausible de entrada, esto es, si tiene una alta probabilidad previa, el primer término del numerador. Como se les dice a los estudiantes de Medicina, si oyes ruido de cascos por la ventana, probablemente se trate de un caballo, no de una cebra. Si ves a un paciente con dolores musculares, es más probable que tenga gripe que kuru (una enfermedad rara hallada en la tribu fore de Nueva Guinea), incluso si los síntomas son compatibles con ambas enfermedades.

En segundo lugar, cree más la idea si es especialmente probable que ocurran las evidencias cuando la idea es verdadera, es decir, si tiene una alta verosimilitud, el segundo término del numerador. Es razonable tomarse en serio la posibilidad de la metahemoglobinemia, también conocida como enfermedad de la piel azul, si se presenta un paciente con la piel azul, o de la fiebre maculosa de las Montañas Rocosas si un paciente de las Montañas Rocosas se presenta con manchas y con fiebre.

Y, en tercer lugar, créela *menos* si las evidencias son habituales, esto es, si tiene una alta probabilidad marginal, el denominador de la fracción. Por eso nos reímos de Irwin el hipocondríaco, convencido de su enfermedad hepática debido a la característica ausencia de malestar. Cierto

es que su carencia de síntomas posee una alta verosimilitud dada la enfermedad, aumentando el numerador, pero nuestra creencia en el autodiagnóstico de Irwin tiene asimismo una probabilidad marginal enorme (ya que la mayoría de la gente no siente malestar alguno la mayor parte del tiempo), inflando el denominador y reduciendo por tanto la probabilidad *a posteriori*.

¿Cómo funciona esto con los números? Regresemos al ejemplo del cáncer. La prevalencia de la enfermedad en la población, 1 %, es cómo establecemos nuestras probabilidades *a priori*: P(hipótesis) = 0,01. La sensibilidad de la prueba es la posibilidad de obtener un resultado positivo dado que el paciente tiene la enfermedad: P(datos | hipótesis) = 0,9. La probabilidad marginal del resultado positivo de una prueba en general es la suma de las probabilidades de un acierto para los pacientes enfermos (el 90 % del 1 %, o 0,009) y de una falsa alarma para los sanos (el 9 % del 99 %, o 0,0891), que podemos redondear como 0,1. Si introducimos las tres cifras en la regla de Bayes, obtenemos 0,01 veces 0,9 dividido entre 0,1, es decir, 0,09.

Así pues, ¿dónde se equivocan los médicos (y, para ser justos, la mayoría de nosotros)? ¿Por qué pensamos que el paciente tiene la enfermedad casi con seguridad, cuando casi con seguridad no la tiene?

OLVIDO DE LA TASA BASE Y HEURÍSTICA DE LA REPRESENTATIVIDAD

Kahneman y Tversky señalaron una importante deficiencia en nuestros razonamientos bayesianos: olvidamos la *tasa base*, que suele ser el mejor dato para calcular la probabilidad *a priori*.[5] En el problema del diagnóstico médico, nuestra mente se dirige al resultado positivo de la prueba (la verosimilitud) y olvidamos lo rara que es la enfermedad en la población (la probabilidad previa).

El dúo fue más allá y sugirió que no ejercitamos el razonamiento bayesiano en absoluto. En su lugar, estimamos la probabilidad de que un caso pertenezca a una categoría en función de lo *representativo* que sea este: de cuánto se asemeje al prototipo o estereotipo de dicha categoría, que representamos mentalmente como una familia borrosa con similitudes

entrecruzadas (capítulo 3). Un paciente con cáncer obtiene normalmente un diagnóstico positivo. Jamás nos planteamos cuán común es el cáncer ni cuán común es un diagnóstico positivo (caballos, cebras, ¿qué más da?). Al igual que la heurística de la disponibilidad del capítulo precedente, la heurística de la representatividad es una regla de oro que el cerebro aplica en lugar de efectuar los cálculos matemáticos.[6]

Tversky y Kahneman demostraron el olvido de la tasa base en el laboratorio contando a los sujetos un accidente con fuga de un taxi por la noche en una ciudad con dos compañías de taxis: Taxi Verde, que posee el 85 % de los vehículos, y Taxi Azul, dueña del 15 % (esas son las tasas base y, por ende, las probabilidades previas). Un testigo presencial identificó el taxi como azul, y los test demostraron que identificaba correctamente los colores de noche el 80 % de las veces (esa es la verosimilitud de los datos, a saber, su testimonio, dado el color real del taxi). ¿Cuál es la probabilidad de que el taxi involucrado en el accidente fuese azul? La respuesta correcta, según la regla de Bayes, es 0,41. La respuesta media fue 0,80, casi el doble de alta. Los encuestados se tomaban demasiado en serio la verosimilitud, prácticamente al pie de la letra, y restaban importancia a la tasa base.[7]

Uno de los síntomas del olvido de la tasa base en el mundo es la hipocondría. ¿A quién de nosotros no le ha preocupado la posibilidad de tener alzhéimer después de un lapso de memoria o un cáncer exótico cuando sentimos un dolor? Otro es el alarmismo médico. Una amiga mía sufrió un momento de pánico cuando un doctor vio moverse nerviosamente a su hija en edad preescolar y sugirió que la niña tenía el síndrome de Tourette. Una vez que se repuso, reflexionó sobre ello como una bayesiana, se percató de que los espasmos son comunes y el Tourette es raro, y volvió a tranquilizarse (aunque le cantó las cuarenta al médico por su analfabetismo estadístico).

El olvido de la tasa base es asimismo un motor del pensamiento en estereotipos. Consideremos a Penelope, una estudiante universitaria descrita por sus amigos como poco práctica y sensible.[8] Ha viajado por Europa y habla francés e italiano con fluidez. Sus planes profesionales son inciertos, pero es una calígrafa talentosa y escribió un soneto para su novio como regalo de cumpleaños. ¿Qué crees que está estudiando Penelope: Psicología

o Historia del Arte? ¡Historia del Arte, por supuesto! ¡Oh!, ¿de veras? ¿Podría ser un poquitín relevante que el 13 % de los universitarios estudian Psicología, pero solo el 0,08 % estudia Historia del Arte, un desequilibrio de ciento cincuenta a uno? Independientemente de dónde veranee o de lo que le regalara a su novio, es poco probable *a priori* que Penelope estudie Historia del Arte. Sin embargo, en nuestra imaginación, es *representativa* de los estudiantes de Historia del Arte, y el estereotipo desplaza las tasas base. Kahneman y Tversky confirmaron esta idea con experimentos en los que pedían a los participantes que considerasen una muestra de setenta abogados y treinta ingenieros (o viceversa), les proporcionaban una breve descripción que encajaba con un estereotipo, como ser un aburrido *nerd*, y les pedían que asignasen una probabilidad al trabajo de esa persona. Los individuos se veían influenciados por el estereotipo; las tasas base entraban por un oído y salían por el otro[9] (este es también el motivo por el que la gente incurre en la falacia de la conjunción del capítulo 1, en la que Linda, la luchadora por la justicia social, tiene más probabilidades de ser una cajera de banco feminista que una cajera de banco; es representativa de una feminista, y la gente se olvida de las tasas base respectivas de las cajeras feministas y las cajeras).

La ceguera hacia las tasas base conduce asimismo a las demandas públicas de lo imposible. ¿Por qué no podemos predecir quién intentará suicidarse? ¿Por qué no disponemos de un sistema de alerta temprana para los tiradores escolares? ¿Por qué no podemos trazar un perfil de los terroristas o de los perpetradores de matanzas y detenerlos preventivamente? La respuesta deriva de la regla de Bayes: una prueba imperfecta de un rasgo raro dará principalmente falsos positivos. El corazón del problema es que solo una proporción ínfima de la población son ladrones, suicidas, terroristas o perpetradores de matanzas (la tasa base). Hasta el día en que los científicos sociales sean capaces de predecir los malos comportamientos con la misma precisión con la que los astrónomos predicen los eclipses, sus mejores pruebas acusarían básicamente a los inocentes e inofensivos.

La conciencia de las tasas base puede ser un don de ecuanimidad cuando reflexionamos sobre nuestra vida. De vez en cuando anhelamos algún resultado extraño: un trabajo, un premio, la admisión en una escuela exclusiva, conquistar el corazón de una persona muy atractiva. Consideramos nuestras eminentes cualificaciones y podemos sentirnos abatidos y resentidos cuando no recibimos la recompensa merecida. Pero, por supuesto, en la carrera participan también otras personas y, por muy superiores que nos creamos, hemos de contar con los demás. Dado que los jueces no son omniscientes, no tenemos ninguna garantía de que vayan a apreciar nuestras virtudes. El recuerdo de las tasas base —el mero número de competidores— puede aliviar el dolor de un rechazo. Por mucho que creamos merecer algo, la tasa base (¿uno de cinco?, ¿uno de diez?, ¿uno de cien?) debería hacer aterrizar nuestras expectativas, y permitirnos calibrar nuestras esperanzas hasta el grado en que quepa esperar razonablemente que nuestra singularidad eleve la probabilidad.

LAS PROBABILIDADES PREVIAS EN CIENCIA Y LA VENGANZA DE LOS LIBROS DE TEXTO

Nuestro olvido de las tasas base es un caso especial de nuestro olvido de las probabilidades *a priori*: el concepto vital, aunque más nebuloso, de cuánto crédito deberíamos dar a una hipótesis antes de examinar las evidencias. Ahora bien, creer en algo antes de examinar las evidencias puede antojarse el epítome de la irracionalidad. ¿No es eso lo que desdeñamos como prejuicio, sesgo, dogma, ortodoxia o ideas preconcebidas? Pero la creencia previa es simplemente el conocimiento falible acumulado a partir de todas nuestras experiencias pretéritas. De hecho, la probabilidad *a posteriori* de una ronda de examen de evidencias puede suministrar la probabilidad *a priori* para la ronda siguiente, un ciclo llamado *actualización bayesiana*. Se trata simplemente de la actitud de alguien que no nació ayer. Para los conocedores falibles en un mundo azaroso, la creencia justificada no puede equipararse al último hecho con el

que nos hemos topado. Como Francis Crick gustaba de decir: «Cualquier teoría capaz de explicar todos los hechos es errónea, porque algunos de los hechos son erróneos».[10]

Este es el motivo por el que resulta razonable ser escéptico con los supuestos milagros, la astrología, la homeopatía, la telepatía y otros fenómenos paranormales, incluso cuando algún testigo presencial o estudio de laboratorio afirme demostrarlo. ¿Por qué no es esto una muestra de dogmatismo y testarudez? Las razones fueron expuestas por ese héroe de la razón llamado David Hume. Hume y Bayes eran contemporáneos y, aunque ninguno de ellos leyó al otro, puede que estuvieran al tanto de sus ideas respectivas a través de un colega mutuo, y el célebre argumento de Hume en contra de los milagros es totalmente bayesiano:[11]

> Nada se considera un milagro si alguna vez ocurre en el curso común de la naturaleza. No es ningún milagro que un hombre, que aparentemente goza de buena salud, muera de repente, porque una muerte semejante, aunque más inusual que cualquier otra, se ha observado frecuentemente que sucede. Pero sí que es un milagro que un hombre muerto regrese a la vida, porque jamás se ha observado tal cosa en ningún tiempo ni país.[12]

En otras palabras, a los milagros tales como la resurrección se les debe asignar una baja probabilidad *a priori*. He aquí la ocurrencia:

> Ningún testimonio es suficiente para establecer un milagro, a menos que el testimonio sea de tal suerte que su falsedad resultase más milagrosa que el hecho que se afana en establecer.[13]

En términos bayesianos, estamos interesados en la probabilidad *a posteriori* de que existan los milagros, dado el testimonio. Contrastémosla con la probabilidad *a posteriori* de que *no* existan los milagros dado el testimonio (en el razonamiento bayesiano, a menudo resulta práctico examinar las *posibilidades*, es decir, la relación entre el crédito de una hipótesis y el crédito de la alternativa, porque ello nos ahorra el engorro de calcular la probabilidad marginal de los datos en el denominador, que es la misma para ambas probabilidades *a posteriori* y se anulan convenientemente). «El hecho que se afana en establecer» es el milagro, con su baja probabilidad previa, que arrastra hacia abajo la posterior. El «testimonio de tal suerte» es la verosimilitud de los datos dado el milagro, y su «falsedad» es la verosimilitud de los datos dado el *no* milagro: la

posibilidad de que el testigo mintiera, percibiera o recordara erróneamente, adornara o transmitiera un relato fantástico que hubiera escuchado a alguien. ¡Habida cuenta de todo cuanto sabemos acerca del comportamiento humano, eso dista de ser milagroso! Esto quiere decir que su verosimilitud es más alta que la probabilidad *a priori* de un milagro. Esa verosimilitud moderadamente elevada incrementa la probabilidad *a posteriori* del no milagro, y reduce las posibilidades totales de un milagro comparado con un no milagro. Otra manera de expresarlo es la siguiente: ¿qué es más probable, que las leyes del universo tal como las entendemos sean falsas o que algún tipo se equivocara en algo?

Una versión más sucinta del argumento bayesiano en contra de las afirmaciones paranormales fue formulada por el astrónomo y divulgador científico Carl Sagan (1934-1996) en la cita que sirve de apertura a este capítulo: «Las afirmaciones extraordinarias requieren evidencias extraordinarias». Una afirmación extraordinaria tiene una baja probabilidad previa bayesiana. Para que su crédito posterior sea más alto que el crédito posterior de su opuesta, la verosimilitud de los datos, dado que la hipótesis es verdadera, ha de ser mucho más alta que la verosimilitud de los datos cuando la hipótesis es falsa. En otras palabras, las evidencias deben ser extraordinarias.

Un fallo del razonamiento bayesiano entre los científicos mismos contribuyó a la crisis de replicabilidad que conocimos en el capítulo 4. El asunto causó un gran revuelo en 2011, cuando el eminente psicólogo social Daryl Bem publicó los resultados de nueve experimentos en la prestigiosa *Journal of Personality and Social Psychology*, que afirmaban demostrar que los participantes predecían exitosamente (a una tasa superior al nivel del azar) sucesos aleatorios antes de que estos tuviesen lugar, tales como cuál de dos cortinas en la pantalla de un ordenador ocultaba una imagen erótica antes de que el ordenador hubiera seleccionado dónde colocarla.[14] No era de extrañar que los efectos no lograran replicarse, pero ese era un resultado inevitable dada la infinitesimal probabilidad previa de que un psicólogo social hubiera refutado las leyes de la física mostrando pornografía a unos estudiantes universitarios. Cuando le planteé esa cuestión a un colega psicólogo social, este replicó: «¡Tal vez Pinker no

comprenda las leyes de la física!». Pero los físicos auténticos, como Sean Carroll en su libro *The Big Picture (El gran cuadro)*, han explicado por qué las leyes de la física descartan realmente la precognición y otras formas de percepción extrasensorial.[15]

El embrollo de Bem planteó un interrogante incómodo. Si un psicólogo eminente podía publicar una tesis ridícula en una revista prestigiosa empleando métodos punteros sometidos a una rigurosa revisión por pares, ¿qué dice todo ello acerca de nuestros estándares de prestigio, eminencia, rigor y del estado de la ciencia? Como hemos visto, una de las respuestas es el peligro de la probabilidad *post hoc*: los científicos habían subestimado los daños que podía causar el fisgoneo de datos y otras prácticas de investigación cuestionables. Pero otra es el desafío al razonamiento bayesiano.

Lo cierto es que la mayoría de los hallazgos en psicología son replicables. Como muchos profesores de Psicología, cada año yo llevo a cabo demostraciones de experimentos clásicos sobre la memoria, la percepción y el juicio ante los alumnos de mis cursos introductorios y de laboratorio, y obtengo los mismos resultados también año tras año. No has oído hablar de esos descubrimientos reproducibles porque no son nada sorprendentes: los individuos recuerdan los ítems del final de una lista mejor que los del medio, o tardan más en rotar mentalmente una letra invertida que una letra tumbada. Los fallos notorios de replicación se producen en estudios que atraían la atención porque sus hallazgos eran sumamente contraintuitivos. Sostener una taza caliente te vuelve más amigable («cálido», ¿lo pillas?). Ver logos de comida rápida te vuelve impaciente. Sujetar un bolígrafo entre los dientes hace que los dibujos animados te parezcan más divertidos, porque obliga a tus labios a esbozar una sonrisa. Las personas a las que se pide que mientan por escrito tienen sentimientos positivos hacia el jabón de manos; los individuos a quienes se le pide que mientan en voz alta tienen sentimientos positivos hacia el enjuague bucal.[16] Cualquier lector de divulgación científica tendrá noticia de otros ingeniosos hallazgos que resultaron aptos para la publicación satírica *Journal of Irreproducible Results [Revista de Resultados Irreproducibles]*.

La razón por la que esos estudios eran presas fáciles para la política de replicabilidad es que tenían bajas probabilidades previas bayesianas. No tan bajas como la percepción extrasensorial, desde luego, pero sería extraordinario descubrir que el estado de ánimo y el comportamiento pudieran zarandearse fácilmente mediante manipulaciones triviales del entorno. Después de todo, industrias enteras de la persuasión y la psicoterapia intentan hacer exactamente eso con grandes costes y con un éxito solo modesto.[17] Era lo extraordinario de los hallazgos lo que les granjeaba un lugar en las secciones de ciencia de los periódicos y en los festivales de ideas llamativas, y por este motivo, siguiendo los parámetros bayesianos, deberíamos exigir evidencias extraordinarias antes de creerlos. De hecho, el sesgo hacia los descubrimientos raros puede convertir el periodismo científico en un surtidor de errores a mansalva. Los editores saben que pueden multiplicar el número de lectores con titulares de portada como los siguientes:

«¿Estaba equivocado Darwin?».

«¿Estaba equivocado Einstein?»

«Joven promesa pone patas arriba el panorama científico.»

«Una nueva revolución científica en *x*.»

«Todo lo que sabes sobre *y* es falso.»

El problema estriba en que *sorprendente* es sinónimo de «baja probabilidad *a priori*», asumiendo que nuestros conocimientos científicos acumulativos no sean despreciables. Esto significa que, incluso si la calidad de las evidencias permanece constante, deberíamos dar *menos* crédito a las afirmaciones sorprendentes. Ahora bien, no se trata tan solo de un problema periodístico. El médico John Ioannidis escandalizó a sus colegas y se anticipó a la crisis de replicabilidad con su artículo «Why most published research findings are false» [«Por qué la mayoría de los resultados de las investigaciones publicados son falsos»]. Un gran problema es que muchos de los fenómenos estudiados por los investigadores biomédicos son interesantes y, *a priori*, es poco probable que sean ciertos, por lo que requieren métodos altamente sensibles para evitar falsos positivos, mientras

que muchos hallazgos verdaderos, incluidos los intentos exitosos de replicación y los resultados nulos, se consideran demasiado aburridos para ser publicados.

Por supuesto, esto no quiere decir que la investigación científica sea una pérdida de tiempo. La superstición y la creencia popular tienen un historial aún peor que la imperfecta ciencia y, a la larga, las turbulencias de las disputas científicas acaban generando conocimiento. Como observó en 1978 el físico John Ziman: «La física de los manuales universitarios es verdadera en un 90 %; los contenidos de las principales revistas de investigación en física son falsos en un 90 %».[18] Este es un recordatorio de que el razonamiento bayesiano recomienda no seguir la práctica común de usar *libro de texto* como un insulto y *revolución científica* como un cumplido.

Un respeto saludable por lo aburrido mejoraría asimismo la calidad de los comentarios políticos. En el capítulo 1 vimos que los historiales de muchos célebres pronosticadores son risibles. Una razón poderosa es que sus carreras dependen de atraer la atención con predicciones fascinantes, esto es, con bajas probabilidades *a priori* y, por tanto, asumiendo que carecen del don de la profecía, con bajas probabilidades *a posteriori*. Philip Tetlock ha estudiado a los «superpronosticadores» que en realidad sí que tienen buenos historiales en sus predicciones de resultados económicos y políticos. Un denominador común es que son bayesianos: parten de una probabilidad previa y la actualizan. Cuando se les pide que calculen la probabilidad de un ataque terrorista el próximo año, por ejemplo, en primer lugar estimarán una tasa base acudiendo a Wikipedia y contando el número de atentados en la región en los años precedentes; una práctica que es poco probable que veamos en el próximo artículo de opinión que leamos sobre lo que le espera al mundo.[19]

TASAS BASE PROHIBIDAS Y TABÚ BAYESIANO

El olvido de la tasa base no siempre es un síntoma de la heurística de la representatividad. A veces se persigue activamente. Las *tasas base prohibidas* constituyen el tercero de los tabúes seculares de Tetlock

(capítulo 2), junto con los contrafácticos heréticos y el intercambio tabú.[20]

El marco para las tasas base prohibidas es creado por una ley de las ciencias sociales. Midamos cualquier variable socialmente significativa: resultados de exámenes, intereses profesionales, confianza social, ingresos, tasas de matrimonio, hábitos de vida, índices de diferentes tipos de violencia (delincuencia callejera o de bandas, violencia doméstica, crimen organizado, terrorismo). Desglosemos acto seguido los resultados en función de los sectores demográficos habituales: edad, sexo, raza, religión y etnicidad. Los promedios de los diferentes subgrupos nunca son los mismos, y a veces las diferencias son considerables. Es irrelevante que estas sean fruto de la naturaleza, la cultura, la discriminación, la historia o alguna combinación de estos factores: las diferencias están ahí.

Esto apenas sorprende, pero tiene una consecuencia espeluznante. Supongamos que buscas la predicción más certera posible acerca de las perspectivas futuras de un individuo: cuánto éxito tendrá la persona en la universidad o en el trabajo, cuál es su riesgo crediticio, cuál es la probabilidad de que cometa un delito, se fugue estando bajo fianza, reincida o perpetre un atentado terrorista. Si fueras un buen bayesiano, partirías de la tasa base para la edad, el sexo, la clase, la raza, la etnicidad y la religión de esa persona, adaptándolas en función de las características particulares del individuo en cuestión. En otras palabras, trazarías un perfil. Incurrirías en prejuicios no por ignorancia, odio, supremacía o ningún otro -*ismo* o fobia, sino partiendo de un esfuerzo objetivo por hacer la predicción más exacta.

Por supuesto, a la mayoría de la gente le horroriza esta idea. Tetlock hacía que los participantes pensaran en los ejecutivos de seguros que tenían que establecer las primas para los diferentes barrios en función de sus respectivos historiales de incendios. Eso no les causaba problemas. Ahora bien, cuando los participantes se enteraban de que los barrios diferían asimismo en su composición racial, tenían sus dudas y censuraban al ejecutivo por ser un buen actuario de seguros. Y si ellos mismos habían estado desempeñando el papel de este, y se enteraban de la terrible verdad acerca de las estadísticas de los barrios, intentaban purificarse moralmente participando como voluntarios en una causa antirracista.

¿Supone esto un nuevo ejemplo de irracionalidad humana? ¿Son acaso «racionales» el racismo, el sexismo, la islamofobia, el antisemitismo y demás fanatismos? ¡Por supuesto que no! Los motivos se remontan a la definición de la racionalidad del capítulo 2: el uso del conocimiento para alcanzar un objetivo. Si la predicción actuarial fuera nuestro *único* objetivo, entonces quizá deberíamos utilizar cualquier pizca de información que pudiera brindarnos la probabilidad previa más precisa. Pero ni que decir tiene que ese no es nuestro único objetivo.

Un objetivo superior es la justicia. Es reprobable tratar a un individuo de acuerdo con su raza, sexo o etnicidad; juzgarlo por el color de su piel o la composición de sus cromosomas en lugar de por el contenido de su carácter. Ninguno de nosotros deseamos ser prejuzgados de esa manera y, en virtud de la lógica de la imparcialidad (capítulo 2), hemos de hacer extensivo ese derecho a todo el mundo.

Por otra parte, solo cuando un sistema *se percibe* como justo —cuando los individuos saben que recibirán un trato justo y no serán prejuzgados por características de su biología o de su historia que están fuera de su control— puede granjearse la confianza de sus ciudadanos. ¿Por qué respetar las reglas del juego cuando el sistema te va a penalizar por tu raza, sexo o religión?

Otro objetivo consiste en evitar las profecías autocumplidas. Si un grupo étnico o un sexo ha sido desfavorecido por la opresión en el pasado, puede que se endosen a sus miembros diferentes características medias en el presente. Si esas tasas base se introducen en fórmulas predictivas que determinan su destino en lo sucesivo, perpetuarán esas desventajas. El problema se está agravando ahora que las fórmulas se entierran en redes de aprendizaje profundo con sus capas profundas e indescifrables (capítulo 3). Una sociedad puede desear racionalmente detener ese ciclo de injusticia aunque su precisión predictiva se vea levemente afectada en ese momento.

Finalmente, las políticas son señales. La prohibición del uso de tasas base étnicas, sexuales o raciales entraña un compromiso público con la igualdad y la justicia que reverbera más allá de los algoritmos permitidos en

una burocracia. Proclama que los prejuicios por *cualquier* motivo son impensables, arrojando un oprobio incluyo mayor sobre los prejuicios arraigados en la enemistad y la ignorancia.

Por consiguiente, la prohibición del uso de las tasas base posee un sólido fundamento en la racionalidad. Ahora bien, un teorema es un teorema, y el sacrificio de la precisión actuarial que hacemos de buen grado en el trato de los individuos por las instituciones públicas puede resultar insostenible en otras esferas. Una de esas esferas son los seguros. A menos que una compañía calcule cuidadosamente los riesgos globales de los diferentes grupos, las indemnizaciones excederán las primas y la compañía aseguradora se hundirá. La aseguradora Liberty Mutual discrimina a los varones adolescentes incluyendo en el cálculo de sus primas sus tasas base más elevadas para los accidentes de tráfico, porque, si no lo hiciera, las mujeres adultas estarían subvencionando la temeridad de estos. No obstante, incluso en este ámbito, las compañías de seguros tienen legalmente prohibida la utilización de ciertos criterios en el cálculo de las tarifas, especialmente la raza y, a veces, el género.

Una segunda esfera en la que no podemos prohibir racionalmente las tasas base es la comprensión de los fenómenos sociales. Si la proporción entre los sexos en un campo profesional no es del 50 %, ¿demuestra eso que sus porteros están impidiendo el acceso de las mujeres, o podría haber una diferencia en la tasa base de mujeres que intentan entrar? Si las entidades hipotecarias rechazan en mayor proporción a los solicitantes pertenecientes a minorías, ¿son por ello racistas o, como el hipotético ejecutivo del estudio de Tetlock, podrían estar utilizando las tasas base de impagos de los diferentes barrios, que simplemente se correlacionan con la raza? Los científicos sociales que investigan estas cuestiones son recompensados con frecuencia por sus molestias con acusaciones de racismo y sexismo. Ahora bien, la prohibición a los científicos sociales y a los periodistas de echar un vistazo a las tasas base detendría los esfuerzos por identificar la discriminación actualmente existente y distinguirla de los legados históricos de las diferencias económicas, culturales o legales entre grupos.

Raza, sexo, etnicidad, religión y orientación sexual se han convertido en zonas de conflicto en la vida intelectual, incluso mientras menguaba el fanatismo manifiesto de toda índole.[21] A mi juicio, una razón fundamental es el hecho de no pensar con claridad en las tasas base y exponer cuándo existen buenas razones para prohibirlas y cuándo no.[22] Pero ese es el problema de los tabúes. Al igual que sucede con la instrucción «no pienses en un oso polar», discutir cuándo aplicar un tabú es en sí mismo un tabú.

BAYESIANOS DESPUÉS DE TODO

A pesar de todos nuestros tabúes, olvidos y estereotipos, es un error dar por perdida a nuestra especie como irremediablemente no bayesiana (recordemos que los sans son bayesianos, toda vez que exigen que un rastro sea definitivo antes de inferir que fue dejado por una especie más rara). Gigerenzer ha argüido que, en ocasiones, las personas ordinarias pisan el suelo firme de las matemáticas cuando parecen estar incumpliendo la regla de Bayes.[23] Los propios matemáticos se quejan de que los científicos sociales emplean con frecuencia las fórmulas estadísticas de manera mecánica: introducen los nú meros y suponen que surge la respuesta. En realidad, una fórmula estadística es solo tan buena como los supuestos subyacentes. Los legos pueden ser sensibles a dichos supuestos, y a veces, cuando parecen estar ignorando la regla de Bayes, puede que simplemente estén ejerciendo la cautela que aconsejaría un buen matemático.

De entrada, una probabilidad *a priori* no es lo mismo que una tasa base, aun cuando las tasas base se presentan con frecuencia como la probabilidad previa «correcta» en las pruebas de lápiz y papel. El problema es *qué* tasa base. Supongamos que obtengo un resultado positivo en la prueba del antígeno prostático específico y quiero calcular mi probabilidad posterior de padecer cáncer de próstata. Para la probabilidad previa, ¿debería usar la tasa base del cáncer de próstata en la población? ¿Entre los estadounidenses blancos? ¿Entre los judíos askenazís? ¿Entre los judíos askenazís mayores de sesenta y cinco años? ¿Los judíos askenazís mayores de sesenta y cinco años que hacen ejercicio y no tienen antecedentes familiares? Estas tasas pueden ser muy diferentes. Por supuesto, cuanto más

específica sea la clase de referencia, mejor. Ahora bien, cuanto más específica es la clase de referencia, más pequeña resulta la muestra en la que se basa el cálculo y más ruidosa la estimación. La mejor clase de referencia constaría de las personas *exactamente* iguales a mí, a saber, yo mismo, una clase de un solo miembro que es perfectamente precisa y perfectamente inútil. No tenemos más opción que utilizar el juicio humano para compensar la especificidad con la fiabilidad al elegir una probabilidad previa apropiada, en lugar de aceptar la tasa base para una población entera estipulada en los términos de una prueba.

Otro problema del uso de una tasa base como probabilidad previa es que las tasas base pueden variar, y a veces con rapidez. Hace cuarenta años, en torno a una décima parte de los estudiantes de Veterinaria eran mujeres; hoy en día la proporción se acerca más a nueve décimas partes.[24] En las últimas décadas, a cualquiera que le dieran la tasa base histórica y la introdujera en la regla de Bayes le habría ido peor que si hubiera olvidado por completo la tasa base. En lo que atañe a muchas de las hipótesis que nos interesan, ninguna agencia de mantenimiento de registros ha compilado siquiera las tasas base (¿sabes qué proporción de estudiantes de Veterinaria son judíos? ¿Y zurdos? ¿Y transgénero?). Y huelga decir que la falta de datos sobre las tasas base ha marcado nuestra difícil condición durante la mayor parte de la historia y la prehistoria, cuando se forjaron nuestras intuiciones bayesianas.

Dado que no existe ninguna probabilidad previa «correcta» en un problema bayesiano, el alejamiento de la gente de la tasa base proporcionada por un experimentador no es necesariamente una falacia. Consideremos el problema del taxi, en el que las probabilidades previas eran las proporciones de taxis azules y verdes en la ciudad. Los participantes bien pueden haber pensado que ese simple punto de partida quedaría inundado por diferencias más específicas, como las tasas de accidentes de las compañías, el número de sus taxis que circulaban durante el día y por la noche, y los barrios en los que prestaban servicios. En tal caso, en su ignorancia de esos datos cruciales, pueden haber adoptado por defecto el valor de indiferencia, 50 %. Los estudios de seguimiento

mostraron que los participantes se vuelven mejores bayesianos cuando se les proporcionan tasas base más relevantes en lo que respecta a estar involucrado en un accidente.[25]

Además, una tasa base puede tratarse como una probabilidad previa solamente cuando los ejemplos disponibles *se seleccionan al azar* en esa población. Si se han escogido cuidadosamente en virtud de un rasgo interesante —como la pertenencia a una categoría con una alta probabilidad de alardear de esos datos—, puede ocurrir cualquier cosa. Tomemos las demostraciones que presentaban a los individuos un estereotipo, como Penelope la escritora de sonetos, o el *nerd* en el grupo de abogados e ingenieros, y les pedían que adivinasen su carrera o su profesión. A menos que los encuestados supieran que Penelope había sido seleccionada por sorteo entre el grupo de estudiantes, lo cual resultaría muy extraño, podrían haber sospechado que había sido elegida porque sus características ofrecían pistas reveladoras, lo cual es una cuestión natural (de hecho, esta cuestión se convirtió en un concurso clásico llamado *What's My Line?*, en el que un equipo tenía que adivinar la ocupación de un misterioso invitado, no seleccionado al azar, por supuesto, sino porque su trabajo era muy característico, como portero de un bar, cazador de caza mayor, jugador de los Harlem Globetrotters o el famoso coronel Sanders del Kentucky Fried Chicken). Cuando a los sujetos les restriega en la cara la aleatoriedad de la muestra (como al ver sacar de un tarro la descripción), sus estimaciones se aproximan más a la probabilidad posterior bayesiana correcta.[26]

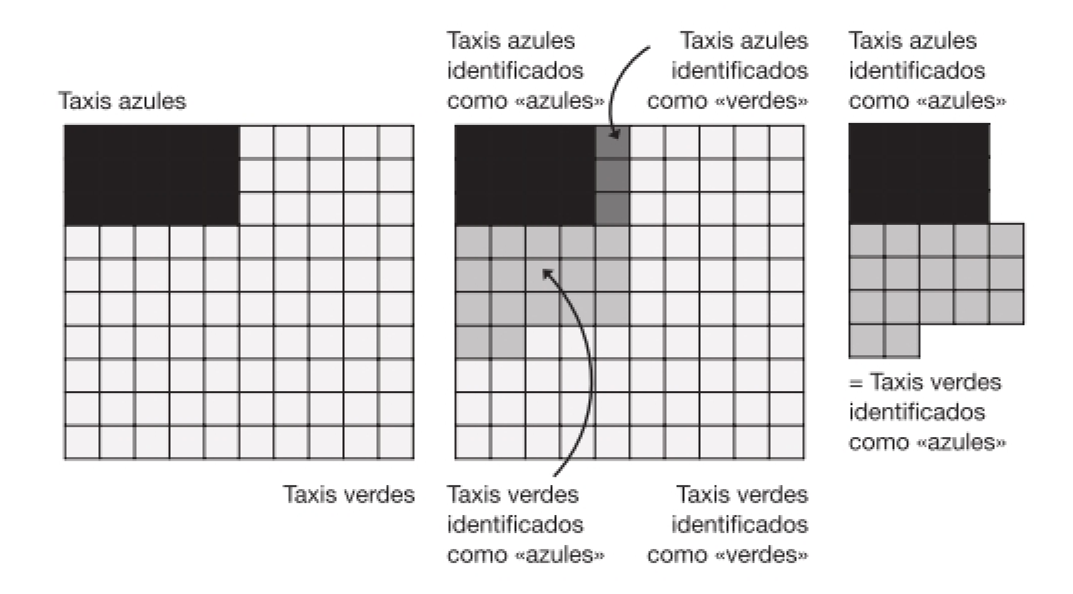
Finalmente, la gente es sensible a la diferencia entre la probabilidad en el sentido de la creencia en un único suceso y en el sentido de frecuencia a largo plazo. Muchos problemas bayesianos plantean el interrogante vagamente místico de la probabilidad de un suceso único: si Irwin tiene kuru, si Penelope estudia Historia del Arte o si el taxi involucrado en un accidente era azul. Al enfrentarse a semejantes problemas, es cierto que los individuos no calculan fácilmente una creencia subjetiva utilizando los números que les dan. Pero, habida cuenta de que incluso los estadísticos están divididos acerca de cuánto sentido tiene eso, quizá puedan ser perdonados. Gigerenzer, junto con Cosmides y Tooby, sostiene que las personas no conectan fracciones decimales con acontecimientos únicos,

porque no es así como la mente humana encuentra la información estadística en el mundo. Experimentamos *sucesos*, no números entre el cero y el uno. Somos perfectamente capaces de efectuar razonamientos bayesianos con estas «frecuencias naturales» y, cuando un problema se reformula en esos términos, podemos recurrir a nuestra intuición para resolverlos.

Regresemos al problema del diagnóstico médico del comienzo del capítulo y traduzcamos esas fracciones metafísicas a frecuencias concretas. Olvidemos el genérico «una mujer»; pensemos en una muestra de mil mujeres. De cada 1.000 mujeres, 10 tienen cáncer de mama (esa es la prevalencia o tasa base). De esas 10 mujeres que tienen cáncer de mama, 9 darán positivo en el test (esa es la sensibilidad al test). De las 990 mujeres sin cáncer de mama, en torno a 89 darán positivo en el test de todos modos (ese es el índice de falsos positivos). Una mujer da positivo. ¿Cuál es la probabilidad de que tenga realmente cáncer de mama? No es tan difícil: 98 de las mujeres dan positivo en total, 9 de ellas tienen cáncer; 9 dividido entre 98 es aproximadamente el 9 %; he aquí nuestra respuesta. Cuando el problema se plantea en estos términos, el 87 % de los médicos lo hace bien (en comparación con alrededor del 15 % con la formulación original), al igual que la mayoría de los niños de diez años.[27]

¿Cómo funciona esta magia? Gigerenzer observa que el concepto de *probabilidad condicional* nos aleja de las cosas contables del mundo. Esas fracciones decimales (el 90 % de positivos verdaderos, el 9 % de falsos positivos, el 91 % de negativos verdaderos, el 10 % de falsos negativos) no suman el cien por cien, de modo que, para calcular la proporción de positivos verdaderos entre todos los positivos (el reto al que nos enfrentamos), tendríamos que efectuar tres multiplicaciones. Las frecuencias naturales, en cambio, nos permiten centrarnos en los positivos y sumarlos: 9 positivos verdaderos más 89 positivos falsos igual a 98 positivos en total, de los cuales los 9 verdaderos suponen el 9 % (lo que deberíamos *hacer* con este conocimiento, dados los costes de actuar o no en función de él, será el tema de los dos próximos capítulos).

Más fácil todavía, podemos usar nuestro cerebro visual de primates y convertir los números en formas. Esto puede tornar asombrosamente intuitivo el razonamiento bayesiano incluso con los acertijos de los libros de texto que están alejados de nuestra experiencia cotidiana, como el problema clásico de los taxis. Visualicemos la flota de taxis de la ciudad como una tabla de 100 cuadrados, uno por taxi (diagrama de la izquierda, abajo). Para representar la tasa base del 15 % de taxis azules, coloreamos 15 cuadrados en la esquina superior izquierda. Para mostrar las verosimilitudes de las cuatro identificaciones posibles por nuestro testigo ocular, que era fiable en un 80 % (diagrama del medio), aclaramos 3 de los cuadrados de taxis azules (el 20 % de los 15 taxis azules que identificaría erróneamente como «verdes») y oscurecemos 17 de los verdes (el 20 % de los 85 taxis verdes que identificaría erróneamente como «azules»). Sabemos que el testigo dijo «azul», así que podemos deshacernos de todos los cuadrados correspondientes a las identificaciones de «verdes», tanto verdaderas como falsas, con lo que nos queda el diagrama de la derecha, que conserva solamente las identificaciones «azules». Ahora es pan comido observar la forma y vislumbrar que la porción más oscura, los taxis que son realmente azules, ocupa un poco menos de la mitad del área total. Si queremos ser exactos, podemos contar: 12 cuadrados de 29, el 41 %. La clave intuitiva tanto para las frecuencias naturales como para las formas visuales es que nos permiten enfocar con el zum los datos disponibles (el resultado positivo del test; las identificaciones «azules»), y clasificarlos en los que son verdaderos y los que son falsos.



Adaptado con el permiso de Presh Talwalkar de Talwalkar 2013

Explotando las intuiciones preexistentes y traduciendo la información a formatos agradables para la mente es posible afinar los razonamientos estadísticos de los individuos. Necesitamos hacerlo. La alfabetización en el riesgo resulta esencial para los médicos, los jueces, los responsables políticos y otras profesiones que tienen nuestras vidas en sus manos. Y dado que todos vivimos en un mundo en el que Dios juega a los dados, el dominio del razonamiento bayesiano y de otras formas de competencia estadística es un bien público que debería constituir una prioridad en la educación. Los principios de la psicología cognitiva sugieren que es preferible trabajar con la racionalidad que posee la gente y mejorarla a descartar a la mayor parte de nuestra especie como crónicamente incapacitada por las falacias y los sesgos.[28] Lo mismo sugieren los principios de la democracia.

CAPÍTULO

6

Riesgo y recompensa
(Elección racional y utilidad esperada)

Todo el mundo se queja de su memoria, pero nadie de su juicio.

FRANÇOIS DE LA ROCHEFOUCAULD

Algunas teorías son antipáticas. Nadie siente mucho afecto por las leyes de la termodinámica, y generaciones de chiflados optimistas han enviado a las oficinas de patentes sus diseños condenados al fracaso de una máquina de movimiento perpetuo. Desde que Darwin propusiera la teoría de la selección natural, los creacionistas se han atragantado con la implicación de que los humanos descendemos de los monos, y los comunitaristas han buscado lagunas en su postulado de que la evolución está impulsada por la competencia.

Una de las teorías más odiadas de nuestro tiempo se conoce en diferentes versiones como elección racional, actor racional, utilidad esperada y *Homo economicus*.[1] Esta última temporada navideña, el programa *CBS This Morning* incluyó un segmento reconfortante sobre un estudio que dejaban miles de carteras llenas de dinero en ciudades del mundo entero y revelaba que la mayoría eran devueltas, especialmente cuando contenían más dinero, recordándonos que, después de todo, los seres humanos somos generosos y honestos. ¿El Grinch de la historia? «Enfoques racionalistas de la economía», que supuestamente predicen que quien se encuentra algo se lo queda.[2]

¿En qué consiste exactamente esta mezquina teoría? Sostiene que, cuando afrontan una decisión arriesgada, los actores racionales deberían elegir la opción que maximice su *utilidad esperada*, esto es, la suma de sus posibles recompensas ponderadas por sus probabilidades respectivas. Fuera de la economía y de unos cuantos rincones de la ciencia política, la teoría se antoja tan adorable como Ebenezer Scrooge, el protagonista del *Cuento de Navidad* de Dickens. Suele interpretarse como una afirmación de que los humanos son, o deberían ser, unos psicópatas egoístas, o unos cerebritos hiperracionales que calculan probabilidades y utilidades antes de decidir si enamorarse o no. Los descubrimientos de los laboratorios de psicología que demuestran que la gente parece violar la teoría se han interpretado como una socavación de las bases de la economía clásica y, por ende, del fundamento de las economías de mercado.[3]

Ahora bien, en su forma original, la teoría de la elección racional es un teorema de las matemáticas, considerado muy hermoso por los aficionados, sin implicaciones directas en lo que atañe a la forma de pensar y elegir de los miembros de nuestra especie. Muchos consideran que ha proporcionado la caracterización más rigurosa de la racionalidad misma, un marco de referencia para medir el juicio humano. Como veremos, esto puede ser impugnado. En ocasiones, cuando la gente se aparta de la teoría, no está claro si está siendo irracional o si los supuestos estándares de la racionalidad son irracionales. Pero, en cualquier caso, la teoría arroja luz sobre los desconcertantes enigmas de la racionalidad y, pese a su procedencia de la matemática pura, puede ser una fuente de profundas lecciones de vida.[4]

La teoría de la elección racional se remonta a los albores de la teoría de la probabilidad y al célebre argumento de Blaise Pascal (1623-1662) sobre por qué deberías creer en Dios. Si crees y no existe, solamente habrás malgastado algunas oraciones, mientras que si no crees y existe, provocarás su ira eterna. Fue formalizada en 1944 por el matemático John von Neumann y el economista Oskar Morgenstern. A diferencia del papa, Von Neumann era considerado realmente un extraterrestre debido a su inteligencia de otro mundo. Inventó asimismo la teoría de juegos (capítulo 8), el ordenador digital, las máquinas autorreplicantes, la lógica cuántica y los componentes claves de las armas nucleares, al tiempo que lograba otras docenas de avances en matemáticas, física e informática.

La elección racional no es una teoría psicológica sobre cómo eligen los seres humanos ni tampoco una teoría normativa acerca de lo que deberían elegir, sino una teoría sobre lo que hace *consistentes* las elecciones con los valores del elector y las unas con las otras. Esto la conecta íntimamente con el concepto de *racionalidad*, que consiste en hacer elecciones consistentes con nuestros objetivos. La búsqueda de Julieta por parte de Romeo es racional, y la búsqueda del imán por parte de las limaduras de hierro no lo es, pues solo Romeo elige qué camino le permite alcanzar su meta (capítulo 2). En el otro extremo de la escala, llamamos *locos* a los individuos cuando hacen cosas evidentemente contrarias a sus intereses, como despilfarrar su dinero en cosas que no desean o correr desnudos con un frío glacial.

La belleza de la teoría estriba en que parte de unos pocos axiomas fáciles de ingerir: requisitos generales aplicables a cualquier tomador de decisiones que estaríamos dispuestos a llamar *racional*. A continuación deduce cómo el decisor tendría que tomar decisiones para permanecer fiel a dichos requisitos. Los axiomas se han agrupado y dividido de varios modos; la versión que presentaré aquí fue formulada por el matemático Leonard Savage y codificada por los psicólogos Reid Hastie y Robyn Dawes.[5]

UNA TEORÍA DE LA ELECCIÓN RACIONAL

El primer axioma puede denominarse *conmensurabilidad*: para cualesquiera opciones A y B, el decisor prefiere A, o prefiere B, o ambas le resultan indiferentes.[6] Esto puede sonar vacuo —¿acaso no son esas justamente las posibilidades lógicas?—, pero requiere que el decisor se comprometa con una de las tres, incluso si se trata de la indiferencia. Esto significa que el decisor nunca recurre a la excusa «no se pueden comparar las peras con las manzanas». Podemos interpretarlo como el requisito de que a un agente racional deben importarle las cosas y ha de preferir unas a otras. No cabe decir lo mismo de los entes no racionales, como las piedras y los vegetales.

El segundo axioma, la *transitividad*, es más interesante. Cuando comparas las opciones de dos en dos, si prefieres A a B y B a C, entonces has de preferir A a C. Resulta fácil ver por qué este es un requisito no negociable: cualquiera que lo viole puede convertirse en una «bomba de dinero». Supongamos que prefieres un iPhone de Apple a un Samsung Galaxy, pero te endosan un Galaxy. Ahora te venderé un elegante iPhone por cien dólares con la permuta. Supongamos que prefieres también un Pixel de Google a un iPhone. ¡Genial! Sin duda canjearías ese cutre iPhone por el superior Pixel más una prima de cien dólares, pongamos por caso. Y supongamos que prefieres un Galaxy a un Pixel: esta es la *intransitividad*. Puedes ver adónde conduce esto. Por cien dólares más una permuta, te venderé el Galaxy. Volverías a estar en el punto de partida, con trescientos dólares menos, y dispuesto a otra ronda de desplume. Cualquiera que sea tu idea de la racionalidad, desde luego no consiste en esto.

El tercer axioma se designa como *cierre*. Con un Dios que juega a los dados y todo lo demás, las elecciones no son siempre entre certezas, como escoger el sabor de un helado, sino que pueden incluir un conjunto de posibilidades con diferentes probabilidades, como escoger un número de lotería. El axioma establece que, siempre que el decisor pueda considerar A y B, ese decisor también puede considerar un número de lotería que ofrece A con una cierta probabilidad p, y B con la probabilidad complementaria $1 - p$.

Dentro de la teoría de la elección racional, aunque el resultado de una opción azarosa no puede predecirse, las probabilidades son fijas, como en un casino. Esto se llama *riesgo* y puede distinguirse de la *incertidumbre*, en la que el decisor no conoce siquiera las probabilidades y puede ocurrir cualquier cosa. En 2002, el secretario de Defensa de Estados Unidos Donald Rumsfeld explicó la distinción: «Hay desconocidos conocidos, lo cual quiere decir que sabemos que hay algunas cosas que no sabemos. Pero también hay desconocidos desconocidos: las cosas que no sabemos que no sabemos». La teoría de la elección racional es una teoría de la toma de decisiones con desconocidos conocidos: con riesgo, no necesariamente con incertidumbre.

Llamaré al cuarto axioma *consolidación*.[7] La vida no solo nos presenta loterías; nos presenta loterías cuyos premios pueden ser a su vez loterías. Una primera cita azarosa, si resulta bien, puede conducir a una segunda cita que comporte todo un nuevo conjunto de riesgos. Este axioma dice simplemente que un decisor enfrentado a una serie de elecciones arriesgadas calcula el riesgo global de acuerdo con las leyes de la probabilidad explicadas en el capítulo 4. Si el primer boleto de lotería tiene una probabilidad de premio de uno sobre diez, y el premio es un segundo boleto con una probabilidad de premio de uno sobre cinco, entonces el decisor lo considera exactamente igual de deseable que un billete con una probabilidad de premio de uno sobre cincuenta (dejaremos al margen el placer adicional que pueda producir una segunda oportunidad de ver los rebotes de las bolas de la lotería o de rascar el billete). Como criterio de

racionalidad, esto parece suficientemente obvio. De la teoría de la probabilidad cabe decir lo mismo que del límite de velocidad y de la gravedad: no es solo una buena idea. Es la ley.

El quinto axioma, la *independencia*, también es interesante. Si prefieres A a B, entonces también prefieres una lotería con A y C como premios a una lotería con B y C como premios (si las probabilidades se mantienen constantes). Es decir, añadir una posibilidad de conseguir C en ambas opciones no debería alterar el hecho de que una sea más deseable que la otra. En otros términos, la manera de *plantear* las alternativas —la forma de presentarlas en un contexto— no debería ser relevante. Comoquiera que la llamemos, una rosa debería tener la misma fragancia. El decisor racional debería centrarse en las opciones mismas y no dejarse confundir por ninguna distracción que acompañe a ambas.

La *independencia de las alternativas irrelevantes*, como se designa la versión genérica de la independencia, es un requisito que figura en muchas teorías de la elección racional.[8] Una versión más simple dice que, si prefieres A a B al elegir entre ambas, deberías seguir prefiriendo A a B al elegir entre ellas y una tercera alternativa C. Cuenta la leyenda que el lógico Sidney Morgenbesser (a quien conocimos en el capítulo 3) estaba en un restaurante y le dieron a elegir tarta de manzana o tarta de arándanos. Poco después de haber escogido la de manzana, regresó la camarera y le indicó que ese día también tenían en el menú tarta de cereza. Como si llevase toda su vida esperando aquel momento, Morgenbesser dijo: «En ese caso, tomaré la de arándanos».[9] Si esto te parece gracioso, quiere decir que aprecias por qué la independencia es un criterio de racionalidad.

El sexto axioma es la *consistencia*: si prefieres A a B, entonces prefieres una apuesta en la que tienes alguna posibilidad de conseguir A, tu primera opción, y no conseguir B, a la certeza de conformarte con B. Media oportunidad es mejor que ninguna.

El último puede llamarse *intercambiabilidad*: la deseabilidad y la probabilidad se compensan.[10] Si el decisor prefiere A a B y prefiere B a C, ha de existir alguna probabilidad de que le resulte indiferente conseguir B con seguridad, su opción intermedia o tener la oportunidad de conseguir A, su primera opción, o conformarse con C. Para que te hagas una idea,

imagínate que partes de una probabilidad del 99 % de conseguir A y solo un 1 % de conseguir C. Esas posibilidades hacen que la apuesta te parezca mucho mejor que conformarte con la segunda mejor opción, B. Ahora considera el otro extremo, una probabilidad del 1 % de conseguir tu primera opción y un 99 % de conseguir tu última opción. Entonces sucede al revés: la opción mediocre segura se impone a la cuasi certeza de tener que conformarte con la peor. Imagínate ahora una secuencia de probabilidades desde la casi certeza de A a la casi certeza de C. Conforme varían gradualmente las posibilidades, ¿crees que aceptarías la apuesta hasta un cierto punto, y te resultaría indiferente apostar o conformarte con B, y después pasarías a la seguridad de B? En tal caso, estás de acuerdo en que la intercambiabilidad es racional.

He aquí la conclusión del teorema. A fin de satisfacer estos criterios de racionalidad, el decisor ha de evaluar el valor de cada resultado en una escala continua de deseabilidad, multiplicarlo por su probabilidad y sumarlos todos para obtener la utilidad esperada de esa opción (en este contexto, *esperada* significa «promedio a largo plazo», no «prevista», y *utilidad* significa «preferible a juicio del decisor», no «útil» ni «práctico»). Los cálculos no tienen por qué ser conscientes ni numéricos; pueden sentirse y combinarse como sentimientos análogos. Entonces, el decisor escogerá la opción con la utilidad esperada más alta. Eso garantizará la racionalidad del decisor en función de los siete criterios. Un decisor racional es un maximizador de la utilidad y viceversa.

Para ser concretos, consideremos una elección entre los juegos de un casino. En los dados, la probabilidad de sacar un «siete» es de una entre seis, en cuyo caso ganarías cuatro dólares; de lo contrario, pierdes un dólar, que es lo que cuesta jugar. Supongamos por ahora que cada dólar es una unidad de utilidad. Entonces, la utilidad esperada de apostar al «siete» en los dados es (1/6 × 4 dólares) + (5/6 × –1 dólar), o –0,17 dólares. Comparemos esto con la ruleta. En la ruleta, la probabilidad de aterrizar en el «siete» es una de 38, en cuyo caso ganarías 35 dólares; de lo contrario perderías tu dólar. Su utilidad esperada es (1/38 × 35 dólares) + 37/38 × –1 dólar), o 0,05 dólares. La utilidad esperada de apostar al «siete» en los dados es más baja que en la ruleta, por lo que nadie te llamaría irracional

por preferir la ruleta (por supuesto, alguien podría llamarte irracional de entrada por el hecho de jugar, ya que el valor esperado de ambas apuestas es negativo, debido a lo que gana la casa, por lo que cuanto más juegas, más pierdes. Ahora bien, si entras en el casino, presumiblemente concedes de entrada una cierta utilidad positiva al glamur de Montecarlo y a la emoción del suspense, lo cual fomenta la utilidad de ambas opciones dentro del territorio de lo positivo, y solamente deja abierta la disyuntiva de a qué jugar).

Con los juegos de azar resulta fácil explicar la teoría de la elección racional, puesto que proporcionan cifras exactas que podemos multiplicar y sumar. Pero en la vida cotidiana nos enfrentamos a un sinfín de elecciones que evaluamos intuitivamente en función de sus utilidades esperadas respectivas. Estoy en una tienda y no recuerdo si hay leche en el frigorífico; ¿debería comprar un litro? Creo que se ha acabado y, en tal caso, si no la compro, me fastidiará de veras tener que comerme los cereales secos mañana por la mañana. Por otra parte, si queda leche en casa y compro más, lo peor que puede ocurrir es que se estropee, pero eso es improbable, e incluso si eso sucede habré perdido solamente un par de dólares. En definitiva, me compensa comprarla. La teoría de la elección racional simplemente especifica el fundamento que hay detrás de esta clase de razonamiento.

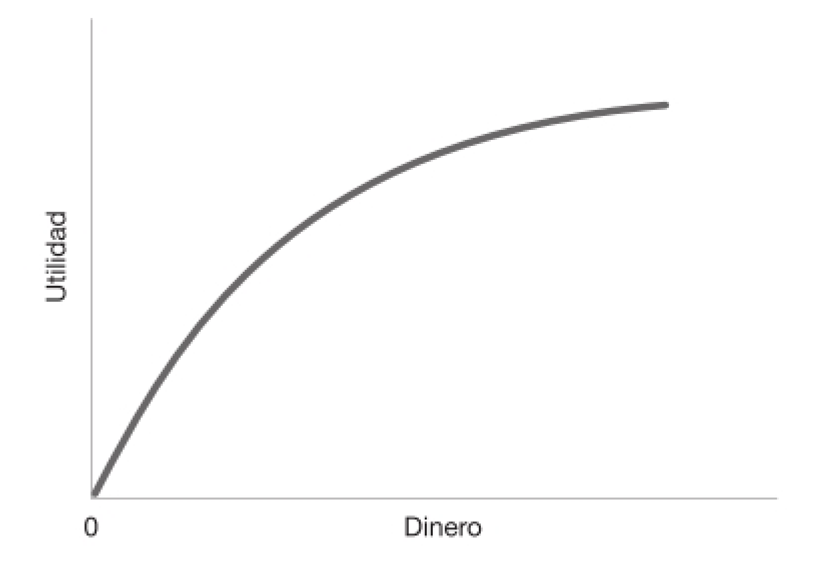
¿Cuán útil es la utilidad?

Resulta tentador pensar que los patrones de preferencias identificados en los axiomas de la racionalidad conciernen a los sentimientos subjetivos de placer y deseo de las personas. Sin embargo, técnicamente hablando, los axiomas tratan al decisor como una caja negra y consideran únicamente sus patrones a la hora de escoger una cosa antes que otra. La escala de utilidad resultante de la teoría es una entidad hipotética reconstruida a partir del patrón de preferencias y recomendada como una manera de preservar la consistencia de dichas preferencias. La teoría protege al decisor de convertirse en una bomba de dinero, un veleta de los postres u otro tipo de

casquivano. Esto significa que la teoría no nos dice tanto cómo actuar de acuerdo con nuestros valores cuanto cómo descubrir nuestros valores observando cómo actuamos.

Esto pone fin a la primera idea errónea sobre la teoría de la elección racional: que retrata a las personas como hedonistas amorales o, lo que es peor, las aconseja que lleguen a serlo. La utilidad no es igual al interés propio; es cualquier escala de valores que un decisor racional maximice. Si los individuos hacen sacrificios por sus hijos y sus amigos, si cuidan de los enfermos y dan limosna a los pobres, si devuelven una cartera llena de dinero, demuestran que el amor, la caridad y la honestidad forman parte de su escala de utilidad. La teoría se limita a ofrecer consejo sobre cómo no malgastarlos.

Por supuesto, al considerarnos a nosotros mismos como tomadores de decisiones, no tenemos por qué tratarnos como cajas negras. La hipotética escala de utilidad debería corresponderse con nuestras sensaciones internas de felicidad, codicia, lujuria, cálido resplandor y demás pasiones. Las cosas cobran interés cuando exploramos la relación partiendo del objeto de deseo más evidente, el dinero. Pueda o no comprar la felicidad, el dinero puede comprar la utilidad, toda vez que la gente cambia cosas por dinero, incluida la caridad. Pero la relación no es lineal; es cóncava. En la jerga, muestra una *utilidad marginal decreciente*.



El significado psicológico es evidente: cien dólares extras aumentan más la felicidad de una persona pobre que de una rica[11] (este es el argumento moral a favor de la redistribución: la transferencia de dinero de los ricos a los pobres incrementa la cantidad de felicidad en el mundo, en igualdad de circunstancias). En la teoría de la elección racional, esta curva no surge en realidad de la fuente evidente, a saber, de preguntar a las personas con diferentes cantidades de dinero cuán felices son, sino de examinar las preferencias de los individuos. ¿Qué preferirías tener: mil dólares con seguridad o una probabilidad del 50 % de ganar dos mil dólares? El valor esperado de ambas cosas es el mismo, pero la mayoría de las personas optan por lo seguro. Esto no significa que incumplan la teoría de la elección racional; significa únicamente que la utilidad no equivale al valor en dólares. La utilidad de dos mil dólares es menos del doble de la utilidad de mil dólares. Afortunadamente para nuestra comprensión, la clasificación que hacen los sujetos de su satisfacción y su elección de apuestas apuntan hacia la misma curva doblada que relaciona el dinero con la utilidad.

Los economistas equiparan la curva de utilidad cóncava con la *aversión al riesgo*. Esto resulta un tanto confuso, porque el término no se refiere a ser un histérico frente a un temerario, sino únicamente a preferir algo seguro a una apuesta con la misma recompensa esperada. No obstante, los conceptos coinciden con frecuencia. La gente contrata seguros a cambio de tranquilidad. Pero lo mismo haría un decisor racional insensible con una curva de utilidad cóncava. El pago de la prima tira de él un poco hacia la izquierda en la escala de dinero, lo cual reduce un poco su felicidad. Ahora bien, si tuviera que sustituir su Tesla no asegurado, su saldo bancario daría un bandazo hacia la izquierda, con un mayor desplome de su felicidad. Nuestro decisor racional opta por la pérdida segura de la prima en lugar de una apuesta con una pérdida mayor, aunque el valor esperado de la pérdida segura (que no ha de confundirse con su utilidad esperada) debe ser un poco más bajo para que la compañía aseguradora obtenga un beneficio.

Desafortunadamente para la teoría, según la misma lógica, la gente jamás apostaría, compraría un boleto de lotería, crearía una empresa o aspiraría al estrellato en lugar de hacerse dentista. Pero, por supuesto,

algunos lo hacen, una paradoja que traía de cabeza a los economistas clásicos. La curva de la utilidad humana no puede ser al mismo tiempo cóncava (lo cual explicaría por qué evitamos el riesgo con los seguros) y convexa (lo que explicaría por qué buscamos el riesgo apostando). Tal vez apostemos por la emoción, al igual que contratamos seguros por la tranquilidad, pero esta apelación a las emociones eleva el nivel de la paradoja: ¿por qué hemos evolucionado con las motivaciones contradictorias de estimularnos y de calmarnos, pagando por ambos privilegios? Quizá seamos irracionales y no haya más que decir al respecto. Tal vez las coristas, las cerezas giratorias de las máquinas tragaperras y otros avíos de los juegos de apuestas constituyan una forma de entretenimiento en la que los jugadores empedernidos estén dispuestos a gastar su dinero. O puede que la curva tenga un segundo cambio de dirección y se dispare hacia arriba en el extremo superior, haciendo que la utilidad esperada de un premio gordo sea más alta que la de un mero incremento en nuestro saldo bancario. Esto podría suceder si los individuos creyeran que el premio los colocaría de un salto en una clase social y un estilo de vida diferentes: la vida del millonario glamuroso y despreocupado, no solo de un miembro más acaudalado de la burguesía. Muchos anuncios de loterías estatales alimentan esa fantasía.

Aunque resulta más fácil reflexionar sobre las implicaciones de la teoría cuando la utilidad se contabiliza en dinero, la lógica es aplicable a cualquier cosa valiosa que podamos situar en una escala. Esto incluye la valoración pública de la vida humana. El dicho falsamente atribuido a Iósif Stalin, en virtud del cual «una muerte es una tragedia, un millón de muertes es una estadística», se equivoca con los números, pero capta nuestra manera de tratar el coste moral de las vidas perdidas en un desastre como una guerra o una pandemia. La curva se dobla, al igual que la de la utilidad del dinero.[12] Un día normal, un atentado terrorista o un incidente de intoxicación alimentaria con una docena de víctimas puede conseguir una amplia cobertura mediática. Ahora bien, en medio de una guerra o de una pandemia, mil vidas perdidas en un día se toman con calma, por más que cada una de esas vidas, a diferencia de la pérdida de un dólar, fuese una persona real, un ser sintiente que amaba y era amado. En *The Better Angels*

of Our Nature (Los ángeles que llevamos dentro), sugerí que nuestro sentido moralmente descaminado de la utilidad marginal decreciente de las vidas humanas es una de las razones por las que las pequeñas guerras pueden llegar a convertirse en catástrofes humanitarias.[13]

¿CUÁN IRRACIONAL ES LA VIOLACIÓN DE LOS AXIOMAS?

Cabría pensar que los axiomas de la elección racional son tan evidentes que cualquier persona normal los respetaría. De hecho, la gente se burla de ellos con frecuencia.

Comencemos con la *conmensurabilidad*. Parece imposible de incumplir: se trata simplemente del requisito de que has de preferir A a B, B a A o que ambas te resulten indiferentes. En el capítulo 2 presenciamos el acto de rebelión, el intercambio tabú.[14] La gente trata ciertas cosas en la vida como sacrosantas, y la mera idea de compararlas se le antoja inmoral. Siente que cualquiera que obedezca el axioma es como el «cínico» Oscar Wilde: alguien que conoce el precio de todo y el valor de nada. ¿Cuánto deberíamos gastar para salvar una especie en peligro de extinción? ¿Y para salvar la vida de una niña que se ha caído a un pozo? ¿Deberíamos equilibrar el presupuesto recortando fondos para la educación, las personas mayores o el medioambiente? Un chiste de otra época comienza con un hombre que pregunta: «¿Te acostarías conmigo por un millón de dólares?».[15] La expresión «la decisión de Sophie», de la desgarradora novela de William Styron, se refiere a que la protagonista tiene que entregar a uno de sus dos hijos para ser gaseado en Auschwitz. En el capítulo 2 vimos cómo rehuir la demanda de comparar entidades sagradas puede ser tanto racional, cuando supone una afirmación de nuestro compromiso con una relación, como irracional, cuando apartamos la mirada de las decisiones dolorosas, pero de hecho las tomamos de manera caprichosa e inconsistente.

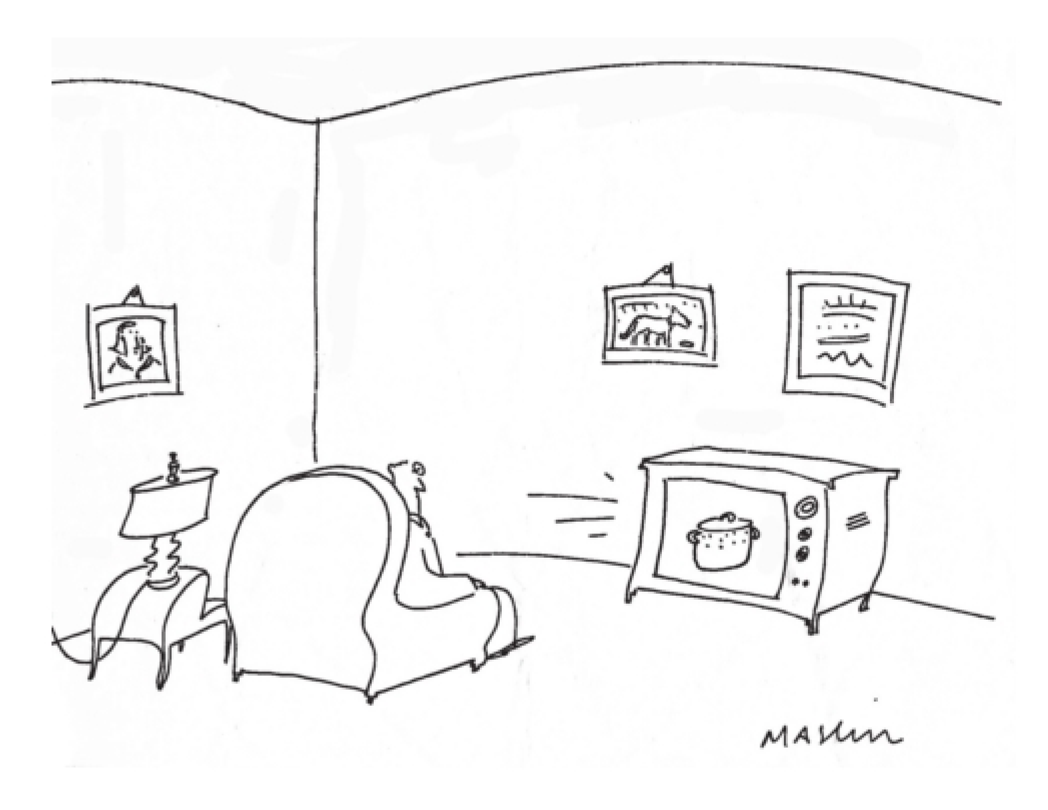
Una familia diferente de violaciones implica un concepto introducido por el psicólogo Herbert Simon, llamado *racionalidad limitada*.[16] Las teorías de la elección racional presuponen un conocedor angelical que dispone de información perfecta, así como de tiempo y memoria ilimitados.

Para los decisores mortales, la incertidumbre respecto de las probabilidades y las recompensas, así como los costes de la obtención y el procesamiento de la información, han de ser tenidos en cuenta en la decisión. Carece de sentido dedicar veinte minutos a buscar un atajo que te ahorrará diez minutos de tiempo de viaje. Los costes no son irrelevantes en modo alguno. El mundo es un jardín de senderos que se bifurcan, en el que todas las decisiones nos conducen a una situación en la que nos enfrentamos a nuevas decisiones, lo cual resulta en una profusión de posibilidades que no podrían domesticarse mediante el axioma de la consolidación. Simon sugería que un decisor de carne y hueso raramente tiene el lujo de optimizar, sino que, en lugar de ello, su opción debe ser *satisfaciente* (*satisfice*, neologismo inglés resultante de la combinación de *satisfy* [«satisfacer»] y *suffice* [«ser suficiente», «bastar»]), es decir, resignarse con la primera alternativa que exceda algún estándar que sea lo bastante bueno. Dados los costes de la información, lo perfecto puede ser enemigo de lo bueno.

Lamentablemente, una regla de decisión que simplifique la vida puede violar los axiomas, incluido el de la transitividad. ¿Incluso la transitividad? ¿Podría yo acaso ganarme la vida encontrando una bomba de dinero humana y vendiéndole las mismas cosas una y otra vez, como Sylvester McMonkey McBean en *The Sneetches (Los sneetches)* del Dr. Seuss, que cobraba reiteradamente tres dólares a los *sneetches* por pegarles una estrella en la barriga y diez dólares por quitársela? («Entonces, cuando se habían gastado hasta el último céntimo de su dinero, el manitas recogía los bártulos y se marchaba».) Aunque la intransitividad es el epítome de la irracionalidad, puede derivar con facilidad de dos características de la racionalidad limitada.

Una de ellas es que no hacemos todas las multiplicaciones y las sumas necesarias para fundir los atributos de un ítem en una masa de utilidad. En vez de ello, podemos considerar sus atributos uno a uno, reduciendo las opciones mediante un proceso de eliminación.[17] Al elegir universidad, podrías descartar de entrada las que no tienen equipo de *lacrosse*, luego las que no tienen Facultad de Medicina, después las que están demasiado lejos de casa, y así sucesivamente.

El otro atajo consiste en que podemos ignorar una pequeña diferencia en los valores de un atributo cuando otros parecen más relevantes. Savage nos pide que pensemos en un turista que no puede decidirse entre visitar París y Roma.[18] Supongamos que, en lugar de ello, le dan a elegir entre visitar París y visitar París más recibir un dólar. París + 1 dólar es incuestionablemente más deseable que París solo. ¡Pero eso *no* significa que París + 1 dólar sea incuestionablemente más deseable que Roma! Tenemos una clase de intransitividad: el turista prefiere A (París + 1 dólar) a B (París), y le resultan indiferentes B y C (Roma), pero no prefiere A a C. El ejemplo de Savage fue redescubierto por un humorista gráfico del *New Yorker*.



«¿Cuánto pagarías por todos los secretos del universo? Espera, no respondas todavía. También recibirías esta olla de vapor de seis litros. ¿Cuánto pagarías ahora?»
Michael Maslin/The New Yorker Collection/The Cartoon Bank

Un decisor que elija mediante un proceso de eliminación puede caer de lleno en la intransitividad.[19] Tversky imagina tres candidatos a un empleo que difieren en sus puntuaciones respectivas en una prueba de aptitud y en

sus años de experiencia:

	Aptitud	Experiencia
Archer	200	6
Baker	300	4
Connor	400	2

Un director de recursos humanos los compara de dos en dos con este criterio: si uno obtiene más de cien puntos más, escoge a ese candidato; en caso contrario, elige al que tenga más experiencia. El director prefiere a Archer a Baker (más experiencia), a Baker a Connor (más experiencia) y a Connor a Archer (más aptitud). Cuando los sujetos experimentales se ponen en el lugar del director, muchos de ellos hacen conjuntos intransitivos de elecciones sin percatarse de ello.

¿Entonces los economistas comportamentales han sido capaces de financiar sus investigaciones utilizando a sus participantes como bombas de dinero? En su mayoría, no. Los individuos entienden el truco, se piensan dos veces sus elecciones y no compran necesariamente algo solo porque lo prefieran momentáneamente.[20] Ahora bien, sin estas reflexiones pausadas del sistema 2, la vulnerabilidad es auténtica. En la vida real, el proceso de toma decisiones mediante la comparación de alternativas de aspecto en aspecto puede dejar al decisor expuesto a las irracionalidades que todos reconocemos en nosotros mismos. A la hora de decidir entre más de dos alternativas, podemos ser influenciados por el último par considerado, o dar vueltas en círculo porque cada alternativa parece preferible a las otras dos en un sentido diferente.[21]

Y los individuos pueden convertirse realmente en bombas de dinero, al menos durante un tiempo, al preferir A a B pero poner un precio más alto a B[22] (les venderías B, se lo cambiarías por A, volverías a comprar A a un precio inferior y repetirías el proceso). ¿Cómo podría incurrir alguien en esta disparatada contradicción? Es fácil: al enfrentarse a dos opciones con el mismo valor esperado, la gente puede preferir aquella que tiene la probabilidad más alta, pero pagar más por la que ofrece una recompensa más alta. Concretamente, consideremos dos billetes para jugar a la ruleta

que tienen el mismo valor esperado, 3,85 dólares, pero con diferentes combinaciones de probabilidades y premios. El billete A te ofrece una probabilidad de 35/36 de ganar cuatro dólares y una probabilidad de 1/36 de perder un dólar. El billete B te ofrece una probabilidad de 11/36 de ganar dieciséis dólares y una probabilidad de 25/36 de perder un dólar y medio.[23] Dada la alternativa, los individuos escogen A. Al preguntarles lo que pagarían por cada uno, ofrecen un precio más alto por B. Es una locura —cuando los sujetos piensan en un precio, se aferran al número mayor que acompaña al signo de dólar y se olvidan de las probabilidades— y el experimentador puede actuar como un arbitrajista y sacar dinero de algunos de ellos. Las desconcertadas víctimas dicen: «No puedo evitarlo», o «Ya sé que es estúpido y que usted se está aprovechando de mí, pero la verdad es que prefiero ese otro».[24] Después de unas cuantas rondas, casi todos espabilan. Parte de la agitación en los mercados financieros reales puede ser provocada porque los inversores ingenuos son influenciados por los riesgos a expensas de las recompensas o viceversa, y los arbitrajistas se abalanzan para explotar las inconsistencias.

¿QUÉ OCURRE CON LA INDEPENDENCIA de las alternativas irrelevantes, con su absurda dependencia del contexto y la formulación? El economista Maurice Allais desveló la siguiente paradoja.[25] ¿Cuál de estos dos billetes de lotería preferirías?

Supercash	100 % de probabilidades de ganar un millón de dólares.	**Powerball**	10 % de probabilidades de ganar 2,5 millones de dólares. 89 % de probabilidades de ganar un millón de dólares.

Aunque el valor esperado del boleto de Powerball es mayor (1,14 millones de dólares), la mayoría de la gente prefiere lo seguro, evitando el 1 % de probabilidades de quedarse sin nada. Esto no viola los axiomas; presumiblemente su curva de utilidad se dobla, provocándoles aversión al riesgo. Veamos ahora cuál de *estos* dos preferirías.

Megabucks	11 % de probabilidades de ganar un millón de dólares.	**LottoUSA**	10 % de probabilidades de ganar 2,5 millones de dólares.

Con esta disyuntiva, los individuos prefieren LottoUSA, lo que sigue la pista de sus respectivos valores esperados (250.000 versus 110.000 dólares). Suena razonable, ¿verdad? Al ponderar la primera opción, el homúnculo de tu cabeza te está diciendo: «La lotería Powerball puede tener un premio mayor, pero, si la escoges, existe la posibilidad de que te quedes sin nada. ¡Te sentirías idiota sabiendo que has perdido un millón de dólares!». Al examinar la segunda opción, dice: «¿Cuál es la diferencia entre un 10 % y un 11 %? En ambos casos tienes alguna posibilidad de ganar; podrías ir por el premio mayor».

Desgraciadamente para la teoría de la elección racional, las preferencias violan el axioma de la independencia. Para apreciar la paradoja, fraccionemos las probabilidades de las dos opciones de la izquierda, manteniendo todo igual excepto la forma de presentarlas:

Supercash	10 % de probabilidades de ganar un millón de dólares. 1 % de probabilidades de ganar un millón de dólares. 89 % de probabilidades de ganar un millón de dólares.	**Powerball**	10 % de probabilidades de ganar 2,5 millones de dólares. 89 % de probabilidades de ganar un millón de dólares
Megabucks	10 % de probabilidades de ganar un millón de dólares. 1 % de probabilidades de ganar un millón de dólares.	**LottoUSA**	10 % de probabilidades de ganar 2,5 millones de dólares.

Vemos ahora que la elección entre Supercash y Powerball es justamente la elección entre Megabucks y LottoUSA con una probabilidad extra del 89 % de ganar un millón de dólares añadida a cada una. Pero esa probabilidad extra te hace invertir tu elección. He añadido tarta de cereza a cada boleto y has cambiado la manzana por los arándanos. Si estás harto de leer acerca de loterías en metálico, Tversky y Kahneman ofrecen un ejemplo no monetario.[26] ¿Preferirías una papeleta para una rifa que

ofreciera una probabilidad del 50 % de ganar un *tour* de tres semanas por Europa o un vale que te garantizase un *tour* de una semana por Inglaterra? La gente prefiere lo seguro. ¿Preferirías una papeleta para una rifa que te ofreciera una probabilidad del 5 % de ganar el *tour* de tres semanas o un boleto con una probabilidad del 10 % de ganar el *tour* por Inglaterra? En este caso la gente se decanta por el viaje más largo.

Psicológicamente está claro lo que sucede. La diferencia entre una probabilidad de cero y una probabilidad del 1 % no es solo una diferencia de un punto porcentual; es la distinción entre la imposibilidad y la posibilidad. Análogamente, la diferencia entre el 99 % y el cien por cien es la distinción entre la posibilidad y la certeza. Ninguna de ambas es conmensurable con las diferencias a lo largo del resto de la escala, como la diferencia entre el 10 % y el 11 %. La posibilidad, por pequeña que sea, da cabida a la esperanza al mirar hacia delante y al arrepentimiento al mirar hacia atrás. Que una elección impulsada por estas emociones sea o no «racional» dependerá de si pensamos que las emociones son respuestas naturales que deberíamos respetar, como comer y mantenernos calientes, o molestias evolutivas que nuestras facultades racionales deberían contener.

Las emociones desencadenadas por la posibilidad y la certeza añaden un ingrediente extra a las elecciones marcadas por las probabilidades, como los seguros y las apuestas, que no pueden explicarse mediante las formas de las curvas de utilidad. Tversky y Kahneman observan que nadie contrataría un seguro probabilístico con primas a una fracción del coste, pero con cobertura solo en ciertos días de la semana, aunque la gente incurre de buen grado en el mismo riesgo global al asegurarse contra algunos peligros, como los incendios, pero no contra otros, como los huracanes.[27] Contratan seguros por la tranquilidad: para tener una cosa menos de la que preocuparse. Prefieren desterrar el miedo por un tipo de desastres del armario de su ansiedad a hacer más seguras sus vidas en general. Esto puede explicar asimismo decisiones sociales tales como prohibir la energía nuclear, con su ínfimo riesgo de un desastre, en lugar de reducir el uso del carbón, con su goteo diario de muchas más muertes. La Ley del Superfondo de Estados Unidos exige la completa eliminación de ciertos contaminantes del medioambiente, aunque la eliminación del último 10 % puede costar

más que el primer 90 %. El juez de la Corte Suprema de Estados Unidos Stephen Breyer comentó a propósito del litigio para obligar a limpiar un lugar de residuos tóxicos: «El informe judicial de cuarenta mil páginas de este esfuerzo de diez años indicaba (y todas las partes parecían estar de acuerdo al respecto) que, sin el gasto adicional, el vertedero de residuos estaba lo suficientemente limpio como para que los niños que jugasen en ese lugar comieran a diario pequeñas cantidades de suciedad durante setenta días al año sin daños significativos [...]. Pero no había ningún niño que comiese suciedad jugando en la zona, pues se trataba de un pantano [...]. Gastar 9,3 millones de dólares para proteger a niños inexistentes de comer suciedad es a lo que me refiero con el problema del último 10 %».[28]

En cierta ocasión le pregunté a un familiar que compraba un boleto de lotería todas las semanas por qué estaba despilfarrando el dinero. Como si fuera un niño lento, me explicó: «No puedes ganar si no juegas». Su respuesta no era necesariamente irracional: puede que haya una ventaja psicológica en el hecho de tener una cartera de perspectivas que incluya la posibilidad de una ganancia inesperada en lugar de maximizar con determinación la utilidad esperada, lo cual garantiza que no pueda producirse dicha ganancia. La lógica se refuerza en un chiste. Un piadoso anciano suplica al Todopoderoso: «Oh, Señor, toda mi vida he obedecido tus leyes. He observado el *sabbat*. He recitado las oraciones. He sido un buen padre y un buen esposo. Solamente te pido una cosa. Quiero ganar la lotería». Los cielos se oscurecen, un rayo de luz atraviesa las nubes y una voz clama: «Veré lo que puedo hacer». El hombre se siente alentado. Pasa un mes, seis meses, un año, pero no le visita la fortuna. En su desesperación grita de nuevo: «Señor Todopoderoso, tú sabes que soy un hombre piadoso. Te he implorado. ¿Por qué me has abandonado?». Los cielos se oscurecen, aparece un rayo de luz y retumba una voz: «Lleguemos a un acuerdo. Compra un boleto».

NO ES SOLO LA FORMULACIÓN de los riesgos lo que puede invertir las elecciones de las personas; también la formulación de las recompensas. Supón que acabas de recibir mil dólares. Ahora has de elegir entre conseguir con seguridad otros quinientos y lanzar una moneda al aire con la

que ganarías otros mil dólares si sale cara. El valor esperado de las dos opciones es el mismo (quinientos dólares), pero a estas alturas has aprendido que la mayoría de la gente siente aversión al riesgo y va a lo seguro. Considera ahora una variación. Imagínate que has recibido dos mil dólares. Ahora has de elegir entre devolver quinientos y lanzar al aire una moneda que te exigiría devolver mil si sale cara. En este caso la mayoría de la gente lanza la moneda. Pero hagamos los cálculos: en términos de dónde acabarías, las opciones son idénticas. La única diferencia está en el punto de partida, que expresa los resultados como una «ganancia» en la primera opción y una «pérdida» en la segunda. Y, con este cambio en la formulación, la aversión al riesgo de los individuos sale por la ventana: ahora *buscan* un riesgo si este les ofrece la esperanza de evitar una pérdida. Kahneman y Tversky concluyen que la gente no siente aversión al riesgo en general, pero sí siente aversión a la *pérdida*: busca el riesgo si este puede evitar una pérdida.[29]

Una vez más, esto no sucede únicamente en los juegos artificiosos. Supongamos que te han diagnosticado un cáncer potencialmente mortal que se puede tratar o bien con cirugía, con algún riesgo de morir en la mesa de operaciones, o bien con radioterapia.[30] A los participantes en el experimento se les indica que, de cada 100 pacientes que optaban por la cirugía, 90 sobrevivían a la operación, 68 seguían vivos un año después y 34 continuaban vivos a los cinco años. En cambio, de cada 100 que optaban por la radioterapia, 100 sobrevivían al tratamiento, 77 seguían con vida al cabo de un año y 22 estaban vivos cinco años después. Menos de un quinto de los sujetos optan por la radioterapia; prefieren la utilidad esperada a largo plazo.

Pero supongamos ahora que las opciones se describen de un modo diferente. De cada 100 pacientes que optaban por la cirugía, 10 morían en la mesa de operaciones, 32 morían después de un año y 66 estaban muertos al cabo de cinco años. De cada 100 que elegían la radiación, no moría ninguno durante el tratamiento, 23 morían en un año y 78 morían en los cinco años siguientes. En este caso, casi la mitad elige la radioterapia. Aceptan una probabilidad global mayor de morir con la garantía de que no morirán inmediatamente por causa del tratamiento. Pero ambos pares de opciones

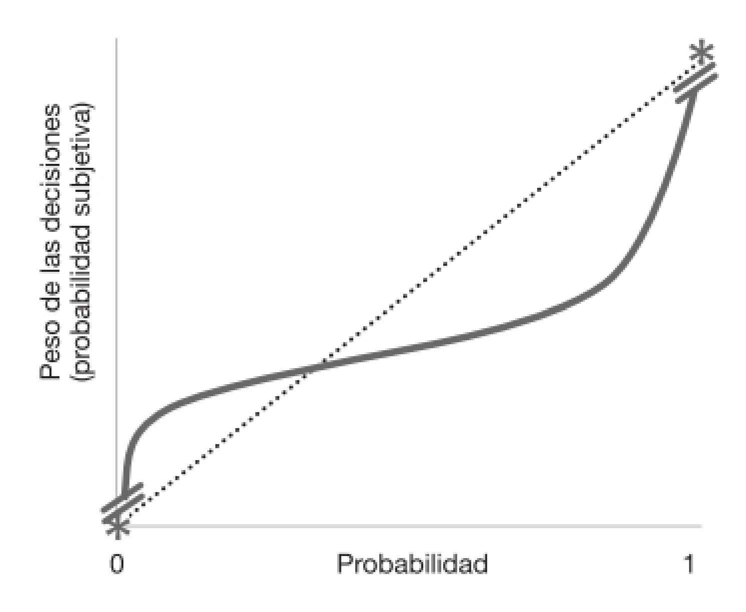
ofrecen las mismas probabilidades. Lo único que cambia es si estas se expresaban como el número de pacientes que viven, percibido como una ganancia, o bien como el número de enfermos que mueren, percibido como una pérdida.

Una vez más, la violación de los axiomas de la racionalidad desborda las elecciones privadas para penetrar en las políticas públi cas. En una espeluznante premonición cuarenta años antes de la COVID-19, Tversky y Kahneman pedían a los sujetos que «imaginasen que Estados Unidos se estuviera preparando para el brote de una inusual enfermedad asiática».[31] Actualizaré su ejemplo. El coronavirus, si se deja sin tratamiento, se espera que mate a 600.000 personas. Se han desarrollado cuatro vacunas y solo una de ellas puede distribuirse a gran escala. Si se escoge Miraculon, se salvarán 200.000 personas. Si se elige Wonderine, hay 1/3 de probabilidades de que se salven 600.000 personas y 2/3 de que no se salve nadie. La mayoría de la gente siente aversión al riesgo y recomienda Miraculon.

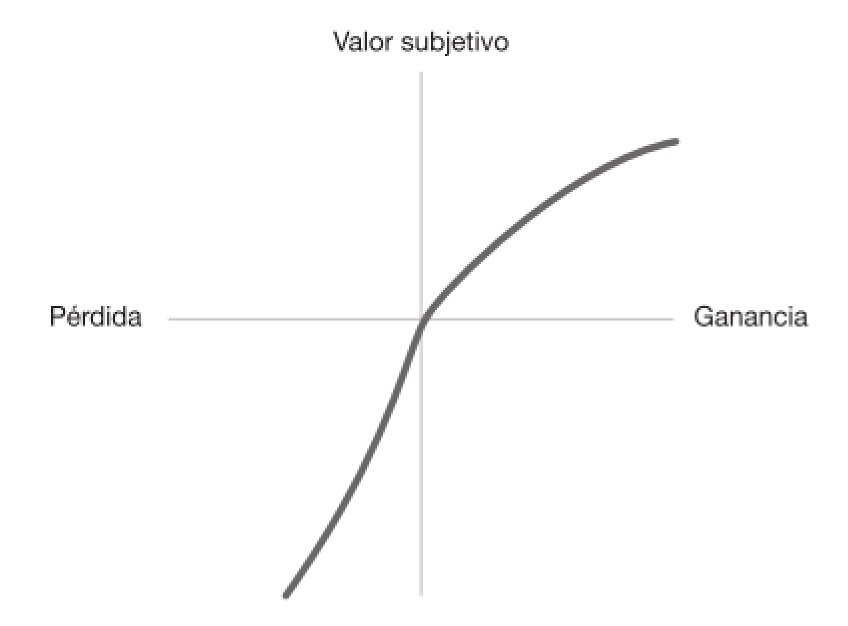
Consideremos ahora las otras dos. Si se elige Regenera, morirán 400.000 personas. Si la escogida es Preventavir, hay una probabilidad de 1/3 de que no muera nadie, y una probabilidad de 2/3 de que mueran 600.000 personas. A estas alturas, has desarrollado un buen ojo para las preguntas capciosas en los experimentos sobre la racionalidad y te has percatado sin duda de que ambas disyuntivas son idénticas, difiriendo solamente en si los efectos se expresan como ganancias (vidas salvadas) o como pérdidas (muertes). Pero la inversión en la formulación ha invertido las preferencias: ahora una mayoría de personas *buscan* el riesgo y optan por Preventavir, que mantiene la esperanza de evitar por completo la pérdida de vidas. No se requiere mucha imaginación para ver cómo podrían explotarse estas formulaciones para manipular a la gente, aunque pueden evitarse con presentaciones cuidadosas de los datos, como mencionando siempre tanto las ganancias como las pérdidas, o presentándolas en forma de gráficos.[32]

Kahneman y Tversky combinaron nuestro sentido deformado de la probabilidad con nuestro veleidoso sentido de las ganancias y las pérdidas en lo que denominan *teoría prospectiva* o *teoría de las perspectivas*.[33] Esta

es una alternativa a la teoría de la elección racional y pretende describir cómo eligen de hecho los individuos, en lugar de prescribir cómo deberían elegir. La gráfica de abajo muestra cómo el «peso de nuestras decisiones», el sentido subjetivo de probabilidad que aplicamos a una elección, está relacionado con la probabilidad objetiva:[34] la curva es pronunciada cerca de cero y de uno (y con una discontinuidad en los límites próximos a esos valores especiales), más o menos objetiva en torno a 0,2, y aplanada en el medio, donde no diferenciamos —pongamos por caso— 0,10 de 0,11.



Un segundo gráfico representa nuestro valor subjetivo.[35] Su eje horizontal está centrado en un punto de referencia móvil, habitualmente el *statu quo*, en lugar de en el cero. El eje no está demarcado en dólares o vidas totales, ni en ningún otro bien apreciado, sino en ganancias o pérdidas relativas con respecto a dicho punto de referencia. Tanto las ganancias como las pérdidas son cóncavas (cada unidad adicional ganada o perdida cuenta menos que las ya obtenidas), pero la pendiente es más pronunciada en el lado negativo; el dolor provocado por una pérdida duplica con creces el placer de la ganancia equivalente.



Por supuesto, la mera representación de los fenómenos como curvas no basta para explicarlos. Pero podemos comprender estas violaciones de los axiomas racionales. La certeza y la imposibilidad son epistemológicamente muy diferentes de las probabilidades muy altas y muy bajas. Por este motivo, en este libro la lógica ocupa un capítulo separado de la teoría de la probabilidad. («P O Q; NO P; por tanto Q» no es solamente un enunciado con una probabilidad muy alta; es una verdad lógica). Por eso las oficinas de patentes devuelven sin abrir las solicitudes de las máquinas de movimiento perpetuo, en lugar de considerar la posibilidad de que algún genio haya solucionado de una vez por todas nuestros problemas energéticos. Benjamin Franklin estaba en lo cierto al menos en la primera mitad de su afirmación de que nada es seguro salvo la muerte y los impuestos. En cambio, las probabilidades intermedias están expuestas a conjeturas, al menos fuera de los casinos. Son estimaciones con márgenes de error, a veces considerables. En el mundo real no resulta descabellado tratar con reservas la diferencia entre una probabilidad de 0,10 y una probabilidad de 0,11.

También la asimetría entre ganancias y pérdidas deviene más explicable cuando descendemos de las matemáticas a la vida real. Nuestra existencia depende de una precaria burbuja de probabilidades, en la que el dolor y la muerte están solo a un paso en falso de distancia. Como me

preguntó Tversky en cierta ocasión cuando éramos colegas: «¿Cuántas cosas podrían sucederte hoy para mejorar mucho tu situación? ¿Cuántas cosas podrían sucederte hoy para *empeorar* mucho tu situación? La segunda lista es interminable». No cabe duda de que estamos más atentos a lo que tenemos que perder, y nos arriesgamos para evitar caídas precipitadas en nuestro bienestar.[36] Y, en el polo negativo, la muerte no es solo un auténtico fastidio. Es el final de la partida, sin posibilidad de volver a jugar, una singularidad que torna irrelevantes todos los cálculos de utilidad.

Este es asimismo el motivo por el que la gente puede violar todavía otro axioma: la intercambiabilidad. Si yo prefiero una cerveza a un dólar y un dólar a la muerte, eso no significa que, con las probabilidades adecuadas, pagaría un dólar para apostar mi vida por una cerveza.

¿O sí?

¿Elecciones racionales después de todo?

En la ciencia cognitiva y en la economía conductual, mostrar todas las formas en las que pueden violarse los axiomas de la elección racional se ha convertido en una especie de deporte (y no solo un deporte: cinco Premios Nobel han sido concedidos a descubridores de estas violaciones).[37] Parte de la diversión consiste en mostrar lo irracionales que somos los humanos; el resto en mostrar lo malos psicólogos que son los economistas y teóricos de la decisión clásicos. A Gigerenzer le encanta contar una historia auténtica acerca de una conversación entre dos teóricos de la decisión, uno de los cuales estaba atormentado por si debía aceptar una tentadora oferta de trabajo en otra universidad.[38] Su colega le dijo: «¿Por qué no pones por escrito las utilidades de permanecer donde estás frente a aceptar el empleo, las multiplicas por sus probabilidades respectivas y escoges la más alta de las dos? Después de todo, eso es lo que aconsejas en tu trabajo profesional». El primero le contestó con brusquedad: «¡Vamos, esto es serio!».

Von Neumann y Morgenstern pueden merecer ser los últimos en reír. Todos esos tabúes, límites, intransitividades, golpes de timón, arrepentimientos, aversiones y formulaciones muestran meramente que la gente viola los axiomas, no que deba hacerlo. Sin lugar a dudas, en ciertos

casos, como en lo que respecta a la sacralidad de nuestras relaciones y al carácter imponente de la muerte, puede compensarnos de veras no hacer las sumas que la teoría prescribe. No obstante, siempre deseamos preservar la consistencia de nuestras elecciones con nuestros valores. Eso es todo cuanto la teoría de la utilidad esperada puede ofrecernos, y se trata de una consistencia que no deberíamos dar por sentada. Tachamos de estúpidas nuestras decisiones cuando subvierten nuestros valores, y las calificamos de sabias cuando los afirman. Ya hemos visto que algunas violaciones de los axiomas son verdaderamente temerarias, como evitar las duras soluciones sociales de compromiso, perseguir el riesgo cero y ser manipulados por las palabras escogidas. Sospecho que hay una infinidad de decisiones en la vida en las que, si multiplicáramos los riesgos por las recompensas, elegiríamos más sabiamente.

Cuando compras un artilugio, ¿deberías comprar asimismo la extensión de la garantía en la que te insiste el vendedor? En torno a un tercio de los estadounidenses lo hacen, invirtiendo 40.000 millones de dólares al año. Ahora bien, ¿realmente tiene sentido contratar un seguro de vida para tu tostadora? Lo que está en juego es menos que en el seguro del coche o del hogar, con los que la pérdida financiera tendría un impacto en tu bienestar. Si los consumidores pensaran incluso toscamente en el valor esperado, se percatarían de que la extensión de la garantía puede costar casi una cuarta parte del precio del producto, lo cual significa que merecería la pena únicamente si el producto tuviera una probabilidad superior a uno sobre cuatro de averiarse. Un vistazo al *Consumer Reports* demostraría entonces que los aparatos modernos no son ni remotamente tan endebles: menos del 7 % de los televisores, por ejemplo, precisan alguna clase de reparación.[39] O consideremos los gastos deducibles del seguro del hogar. ¿Deberías pagar cien dólares adicionales al año para reducir los gastos de tu bolsillo en el caso de dar un parte al seguro por un valor de mil a quinientos dólares? Mucha gente lo hace, pero eso solo tiene sentido si esperas dar un parte cada cinco años. La tasa media de partes para los seguros del hogar se sitúa de hecho en torno a uno cada *veinte* años, lo cual significa que la gente está pagando cien dólares por veinticinco en valor esperado (el 5 % de quinientos dólares).[40]

Con consecuencias de mucho mayor calado, sopesar riesgos y recompensas puede informar asimismo las decisiones médicas. Tanto los médicos como los pacientes tienen tendencia a pensar en términos de propensiones: el cribado de cáncer es bueno porque detecta cánceres, y la cirugía del cáncer es buena porque puede extirparlos. Pero pensar en los costes y los beneficios sopesados en función de sus probabilidades puede convertir lo bueno en malo. Por cada mil mujeres que se someten a ecografías anuales para detectar el cáncer de ovarios, seis son diagnosticadas correctamente con la enfermedad, comparadas con cinco de cada mil mujeres no examinadas, y el número de muertes en ambos grupos es el mismo: tres. Hasta ahí los beneficios. ¿Y qué hay de los costes? ¿De las mil mujeres examinadas, otras noventa y cuatro reciben falsas alarmas aterradoras, treinta y una de las cuales sufren innecesariamente la extirpación de sus ovarios, de las cuales cinco sufren además graves complicaciones. Huelga decir que el número de falsas alarmas y cirugías innecesarias entre las mujeres no examinadas es cero. No hacen falta muchas matemáticas para demostrar que la utilidad esperada del cribado de cáncer de ovario es negativa.[41] Otro tanto sucede con los hombres en lo tocante al cribado para el cáncer de próstata con la prueba de antígeno prostático específico (yo opto por no hacerlo). Estos son casos sencillos; analizaremos con más profundidad cómo comparar los costes y los beneficios de los aciertos y las falsas alarmas en el próximo capítulo.

Incluso cuando no se dispone de cifras exactas, es aconsejable multiplicar mentalmente las probabilidades por los resultados. ¿Cuántas personas han arruinado su vida por apostar con una gran probabilidad de una pequeña ganancia y una pequeña probabilidad de una pérdida catastrófica, saltándose la legalidad por un poco de dinero extra que no necesitaban o arriesgando su reputación y su tranquilidad por una insignificante aventura? Pasando de las pérdidas a las ganancias, ¿cuántos solteros solitarios se privan de la pequeña probabilidad de una vida de felicidad con un alma gemela porque piensan únicamente en la gran probabilidad de un café tedioso con una persona aburrida?

En cuanto a jugarse la vida, ¿alguna vez has ahorrado un minuto en la carretera conduciendo por encima del límite de velocidad, o has satisfecho tu impaciencia comprobando tus mensajes mientras cruzabas la calle? Si sopesaras esos beneficios frente a la probabilidad de un accidente multiplicada por el precio que pongas a tu vida, ¿hacia qué lado se inclinaría la balanza? Y, si no piensas de esta manera, ¿puedes llamarte racional?

CAPÍTULO

7

Aciertos y falsas alarmas
(Detección de señales y teoría estadística de la decisión)

El gato que se sienta sobre la tapa de una estufa caliente [...] nunca volverá a sentarse sobre la tapa de una estufa caliente, y eso está bien; pero tampoco volverá a sentarse en una fría.

MARK TWAIN[1]

La racionalidad requiere que distingamos lo que es verdadero de lo que queremos que sea verdadero: que no enterremos la cabeza en la arena, ni construyamos castillos en el aire ni decidamos que las uvas que están fuera de nuestro alcance están verdes. Las tentaciones de la ilusión y del pensamiento mágico están siempre con nosotros, porque nuestra fortuna depende del estado del mundo, que nunca podemos conocer con certeza. Para mantener nuestra iniciativa y protegernos contra la adopción de medidas dolorosas que pueden revelarse innecesarias, tendemos a ver aquello que deseamos ver y a ignorar el resto. Nos tambaleamos en el borde de la báscula del baño de una forma que minimiza nuestro peso, dejamos para más tarde una prueba médica que puede dar un resultado desagradable y tratamos de creer que la naturaleza humana es infinitamente maleable.

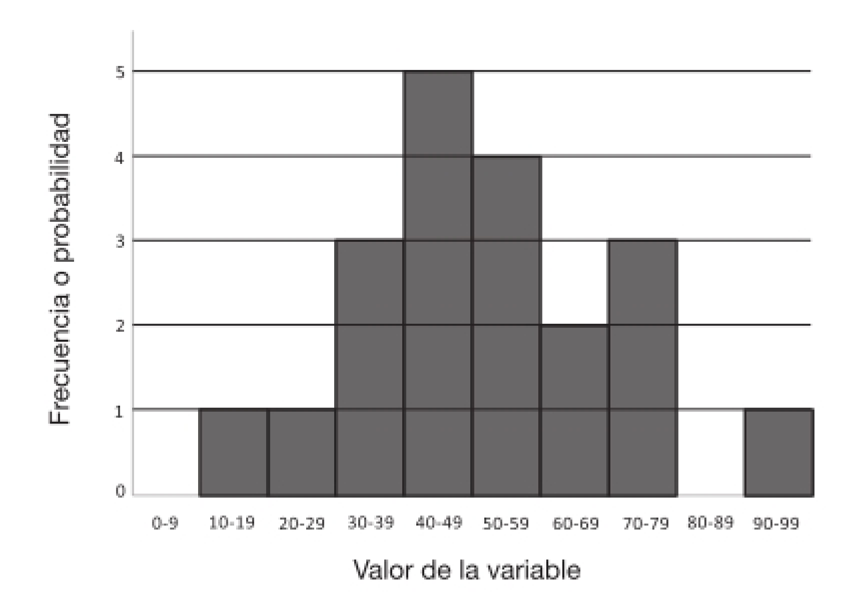
Existe una manera más racional de reconciliar nuestra ignorancia con nuestros deseos: la herramienta de la razón denominada *teoría de la detección de señales* o *teoría estadística de la decisión*. Esta combina las dos grandes ideas de los capítulos precedentes: estimar la probabilidad de que algo sea verdadero acerca del mundo (razonamiento bayesiano) y decidir qué hacer al respecto sopesando sus costes y beneficios esperados (elección racional).[2]

El reto de la detección de señales es si tratar un indicador como una señal genuina del mundo o como ruido en nuestra imperfecta percepción de este. Se trata de un dilema recurrente en la vida. Un centinela ve un punto en la pantalla de un radar: ¿estamos siendo atacados por bombarderos nucleares o se trata de una bandada de gaviotas? Un radiólogo ve una mancha en un escáner: ¿tiene cáncer el paciente o es un quiste inofensivo? Un jurado oye el testimonio del testigo presencial en un juicio: ¿es culpable el acusado o son incorrectos los recuerdos del testigo? Nos encontramos con una persona que nos resulta vagamente familiar: ¿la conocíamos ya o se trata de un *déjà vu* espontáneo? Un grupo de pacientes mejora después de tomar un fármaco: ¿hizo algo el medicamento o fue un efecto placebo?

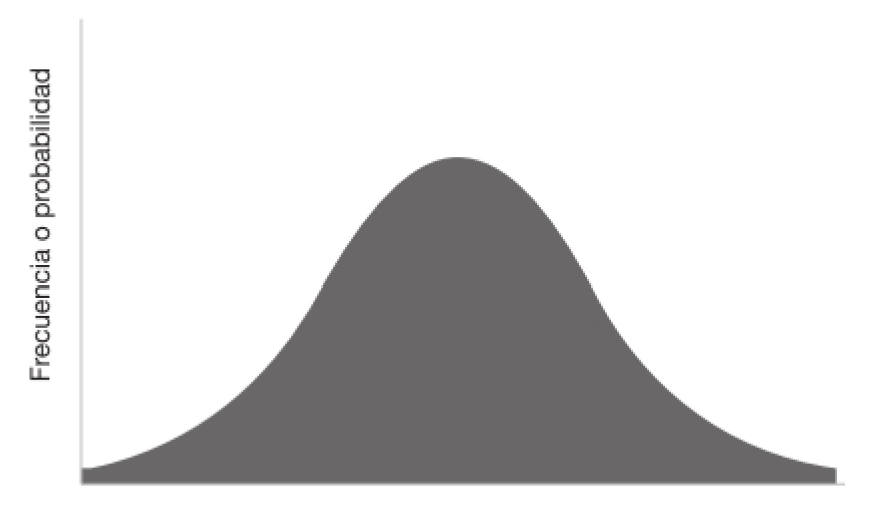
El resultado de la teoría estadística de la decisión no es un grado de creencia, sino una decisión ejecutable: someterse o no a cirugía, condenar o absolver. Al decantarnos por una opción u otra no estamos decidiendo qué creer acerca del estado del mundo. Nos estamos comprometiendo con una acción en espera de sus costes y beneficios probables. Esta herramienta cognitiva nos machaca con la distinción entre lo que es verdadero y lo que hemos de hacer. Reconoce que los diferentes estados del mundo pueden requerir diferentes decisiones arriesgadas, pero muestra que no tenemos por qué engañarnos a nosotros mismos acerca de la realidad para jugar con las probabilidades. Distinguiendo claramente nuestra evaluación del estado del mundo de lo que decidamos hacer al respecto, podemos actuar racionalmente *como si* algo fuese verdadero sin necesidad de *creer* que es verdadero. Como veremos, esto marca una diferencia enorme, pero poco apreciada, en la comprensión del uso de la estadística en la ciencia.

SEÑALES Y RUIDO, SÍES Y NOES

¿Cómo deberíamos pensar acerca de un indicador errático del estado del mundo? Comencemos con el concepto de *distribución estadística*.[3] Supongamos que medimos algo que varía impredeciblemente (una variable aleatoria), como las puntuaciones de un test de introversión de cero a cien. Clasificamos los resultados en intervalos (de cero a nueve, de diez a diecinueve, y así sucesivamente) y contamos el número de personas que caen en cada intervalo. Acto seguido los apilamos en un *histograma*, un gráfico que difiere de los que solemos ver, en el que la variable de interés se representa a lo largo del eje horizontal en lugar del vertical. La dimensión de arriba abajo simplemente apila el número de personas que caen en cada intervalo. He aquí un histograma de las puntuaciones de introversión de veintiuna personas, una por cuadrado.



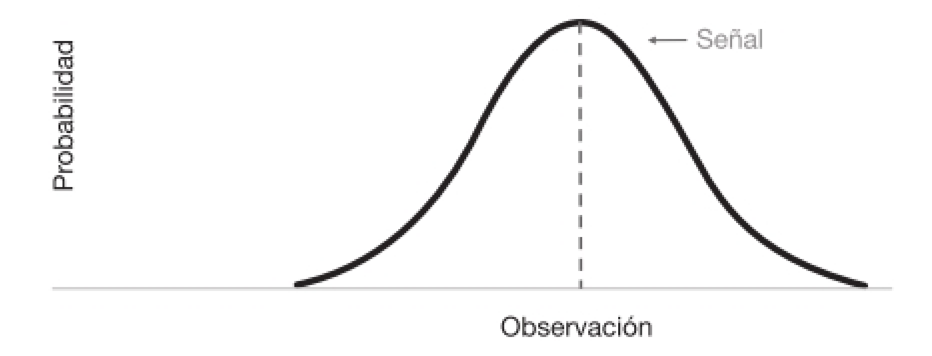
Imaginemos ahora que hemos realizado la prueba a varios *millones* de personas, suficientes para no tener que clasificarlas ya en intervalos, sino que podemos distribuirlas de izquierda a derecha en función de sus puntuaciones originales. Conforme apilamos cada vez más cuadrados y nos vamos alejando, el zigurat se desdibuja convirtiéndose en una suave colina, la familiar curva en forma de campana del gráfico que hay a continuación. Esta contiene muchas observaciones amontonadas en un valor promedio en el medio, y cada vez menos a medida que nos fijamos en los valores que son cada vez más pequeños a la izquierda o cada vez más grandes a la derecha. El modelo matemático más familiar de una curva de campana se denomina *distribución normal* o *gaussiana*.



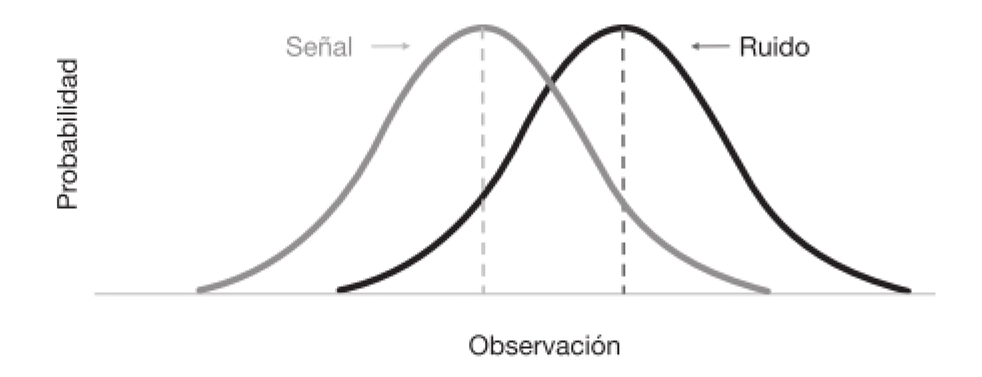
Las curvas en forma de campana son comunes en el mundo, como las puntuaciones de los test de personalidad o inteligencia, las estaturas de hombres y mujeres, y las velocidades de los coches en una carretera. Las curvas de campana no son la única manera de apilar observaciones. Existen asimismo las distribuciones de dos jorobas o bimodales, como el grado relativo de atracción sexual de los hombres hacia las mujeres y hacia los hombres, que tiene un pico grande en un extremo para los heterosexuales y un pico más pequeño en el otro extremo para los homosexuales, con un número aún menor de bisexuales entre medias. Y hay distribuciones de cola gruesa en las que los valores extremos son raros, pero no astronómicamente raros, como las poblaciones de las ciudades, los ingresos de los individuos o el número de visitantes de los sitios web. Muchas de estas distribuciones, como las generadas por *leyes de potencia*, tienen una alta espina dorsal a la izquierda, con muchos valores bajos, y una cola larga y gruesa a la derecha, con un mínimo de valores extremos.[4] Pero las curvas de campana (unimodales, simétricas, de cola delgada) son comunes en el mundo; surgen cada vez que una medida es la suma de un gran número de pequeñas causas, como muchos genes junto con numerosas influencias ambientales.[5]

Pasemos al tema que nos ocupa: las observaciones de si algo ha sucedido o no en el mundo. No podemos adivinar esto a la perfección —no somos Dios—, sino solo mediante nuestras mediciones, como los puntos en la pantalla de un radar provenientes de un avión, o la opacidad de las

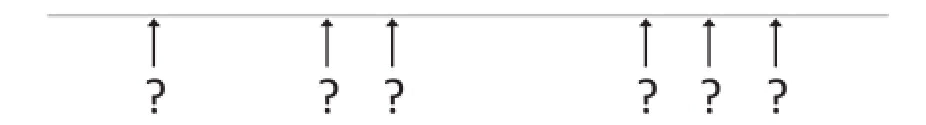
manchas en un escáner producidas por un tumor. Nuestras mediciones no resultan exactamente iguales todas las veces. Antes bien, tienden a distribuirse en una curva de campana, mostrada en el diagrama de abajo. Podemos interpretarlo como una representación de la verosimilitud bayesiana: la probabilidad de una observación dada la presencia de una señal.[6] Por término medio, la observación tiene un valor determinado (la línea discontinua vertical), pero a veces es un poco más alta o más baja.



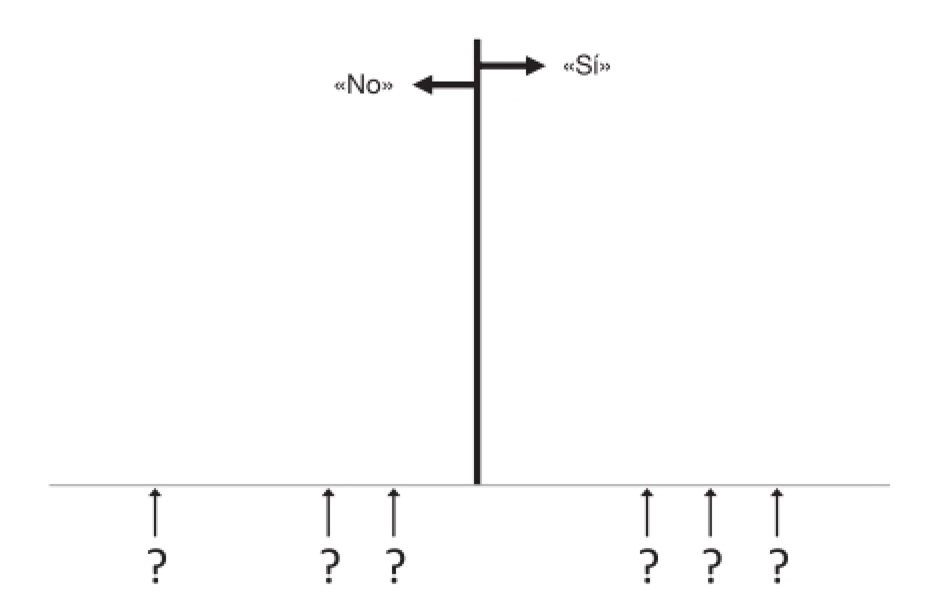
Pero he aquí un giro trágico. Cabría pensar que, cuando no suceda nada en el mundo (ningún bombardero, ningún tumor), obtendremos una medición de cero. Desafortunadamente, eso no ocurre jamás. Nuestras mediciones están siempre contaminadas por el ruido (estática de radio, perturbaciones como una bandada de aves, quistes inofensivos que aparecen en el escáner), y estos también variarán de una medición a otra, cayendo en su propia curva de campana. Más lamentable todavía es el hecho de que el rango superior de las mediciones causadas por el ruido puede solaparse con el rango inferior de las mediciones producidas por el «objeto» o el «hecho» del mundo:



La tragedia consiste en que solo Dios puede ver el diagrama y saber si una observación procede de una señal o del ruido. Todo cuanto vemos los mortales son nuestras observaciones:

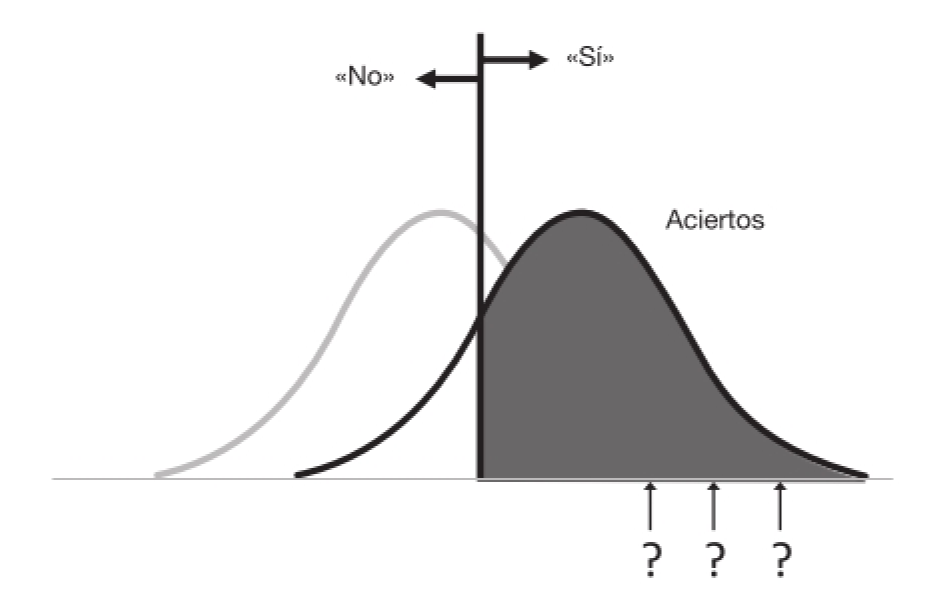


Cuando nos vemos obligados a conjeturar si una observación es una señal (que refleja algo real) o ruido (la confusión de nuestras observaciones), hemos de aplicar un corte. En la jerga de la detección de señales, este se designa como *criterio* o *sesgo de respuesta*, y se simboliza como ß (beta). Si una observación está por encima del criterio, decimos «sí» y actuamos como si fuese una señal (tanto si lo es como si no lo es, cosa que no podemos saber); si está por debajo, decimos «no» y actuamos como si fuese ruido:

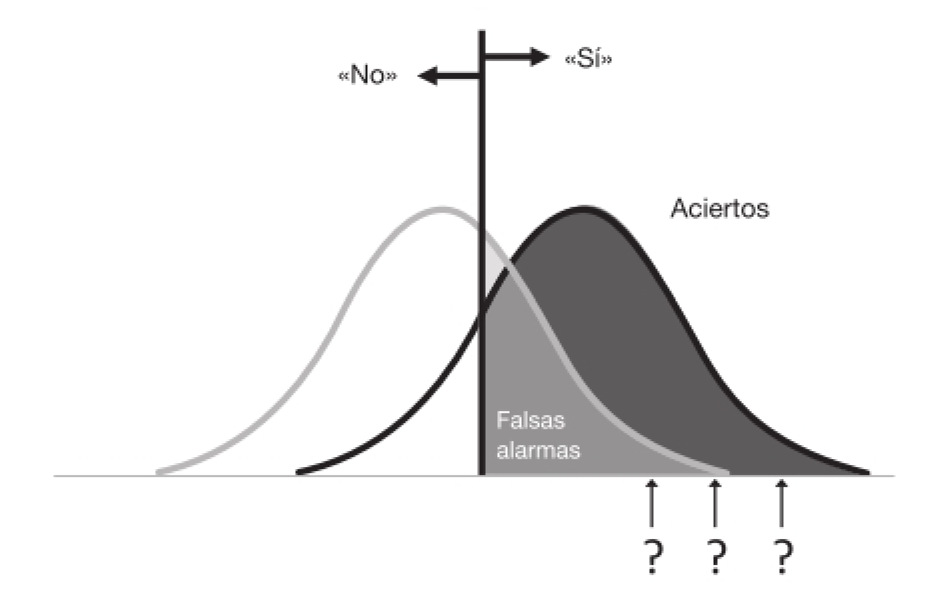


Volvamos a la visión del ojo de Dios y veamos qué tal lo hacemos, por término medio, con este corte. Existen cuatro posibilidades. Cuando decimos «sí» y se trata efectivamente de una señal (el bombardero o el

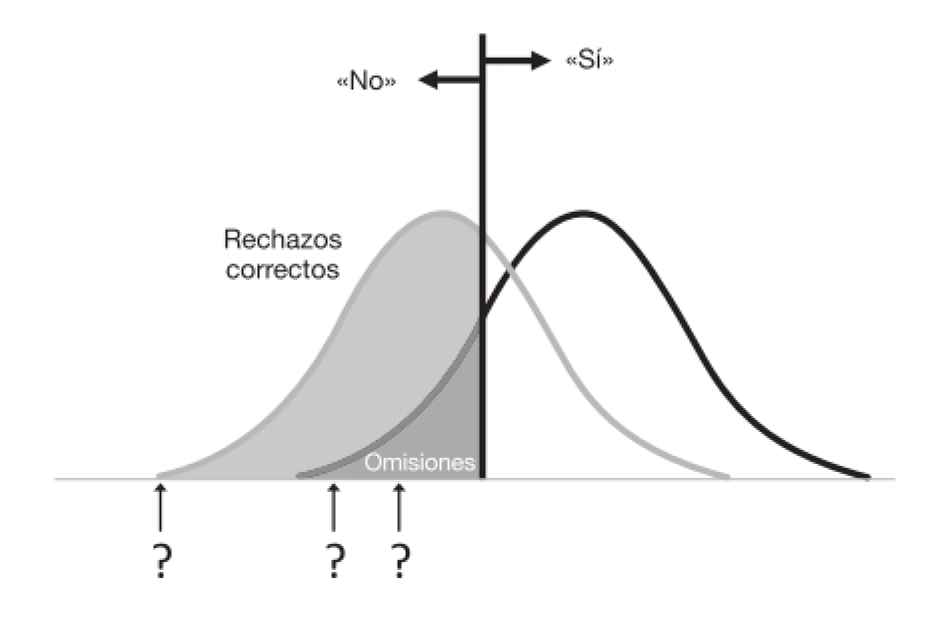
tumor están ahí), hablamos de acierto, y la proporción de señales que identificamos correctamente se representa en la porción de la distribución sombreada en color oscuro:



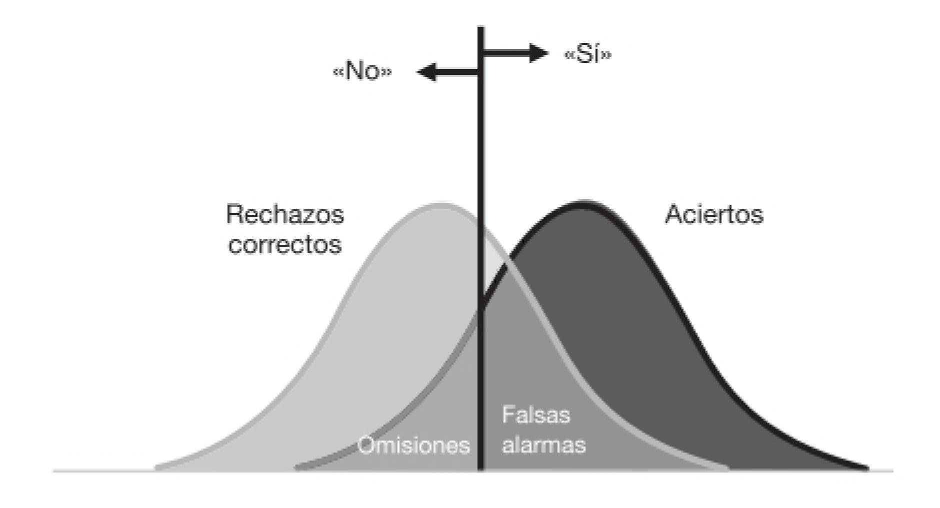
¿Y si no fuese más que ruido? Cuando decimos «sí» a nada, hablamos de una falsa alarma, y la proporción de nadas en la que nos precipitamos se representa como la porción con la tonalidad media de gris:



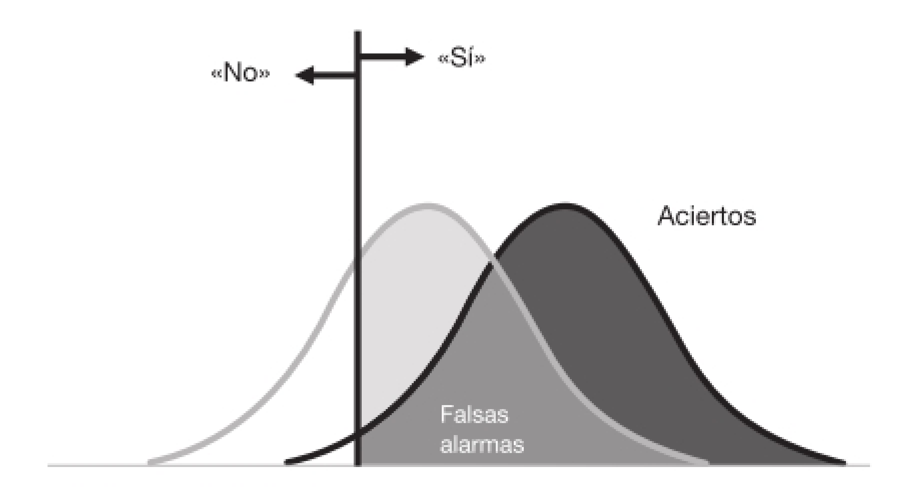
¿Qué ocurre con las ocasiones en las que la observación cae *por debajo de* nuestro criterio y decimos «no»? De nuevo existen dos posibilidades. Cuando está sucediendo realmente algo en el mundo, hablamos de *omisión*. Cuando no hay nada más que ruido, decimos que es un *rechazo correcto*.



He aquí cómo las cuatro posibilidades se reparten el espacio de los sucesos:



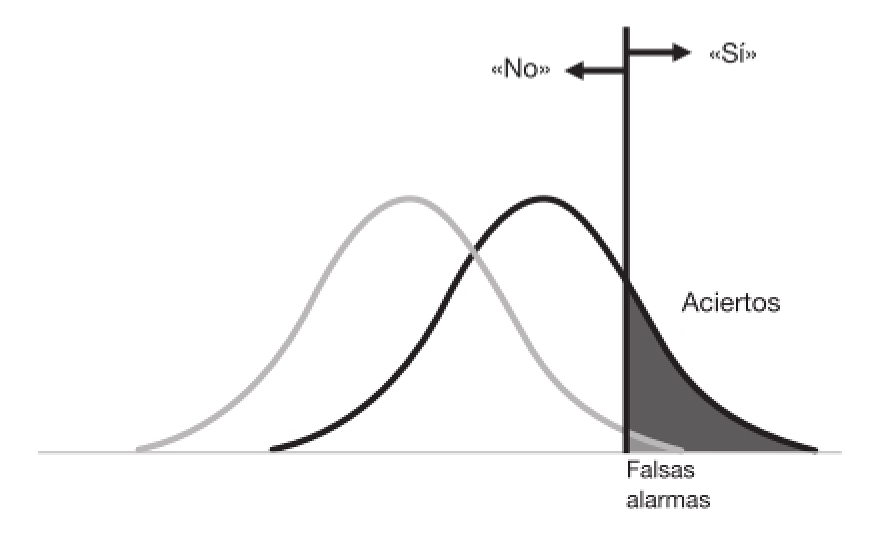
Dado que decimos «sí» o «no» cada vez, las proporciones de aciertos y omisiones cuando existe una señal real (montón derecho) deben sumar el cien por cien. Otro tanto sucede con las proporciones de falsas alarmas y rechazos correctos cuando no hay nada más que ruido (montón izquierdo). Si bajáramos nuestro criterio en el lado izquierdo, disparando con más facilidad, o lo subiéramos en el lado derecho, disparando con más comedimiento, estaríamos cambiando aciertos por omisiones, o falsas alarmas por rechazos correctos, por pura aritmética. Dado que ambas curvas se solapan, resulta menos evidente que *también* estaríamos cambiando aciertos por falsas alarmas (cuando decimos «sí») y omisiones por rechazos correctos (cuando decimos «no»). Examinemos con más detenimiento lo que sucede cuando relajamos el criterio de respuesta, disparando con más alegría o diciendo «sí» con más facilidad:



La buena noticia es que tenemos más aciertos y captamos casi todas las señales. La mala noticia es que tenemos más falsas alarmas, pues decimos que viene el lobo la mayoría de las veces en que no hay nada más que ruido. ¿Y qué sucede si, en lugar de ello, adoptamos un sesgo de respuesta más estricto, que nos lleva a disparar con más comedimiento y a exigir una alta carga de la prueba?

Ahora se vuelven las tornas: apenas disparamos con facilidad con una falsa alarma, lo cual es bueno, pero ignoramos la mayoría de las señales, lo cual es malo. En el caso extremo, si dijéramos descuidadamente «sí» cada

vez, siempre acertaríamos cuando hubiese una señal y siempre nos equivocaríamos cuando hubiera ruido, y al contrario si dijéramos «no» todas las veces.



Esto parece evidente, pero confundir el sesgo de respuesta con la exactitud fijándonos únicamente en las señales o únicamente en el ruido es una falacia sorprendentemente común. Supongamos que un examinador analiza por separado los resultados de los ítems verdaderos y falsos en un test de verdadero o falso. Cree estar viendo si a la gente se le da mejor detectar verdades o rechazar falsedades, pero todo lo que está viendo en realidad es si son la clase de personas a quienes les gusta decir «sí» o «no». Me quedé horrorizado cuando un médico me pasó una prueba de audición que presentaba una serie de pitidos con un volumen creciente, desde lo inaudible hasta lo imposible de ignorar, y me pidió que levantara un dedo cuando empezara a oírlos. No era una prueba de mi audición. Era una prueba de mi impaciencia y mi disposición a jugármela cuando era incapaz de decir sinceramente si estaba oyendo un tono o un acúfeno. La teoría de la detección de señales nos proporciona una serie de formas de hacer esto correctamente, entre las que figuran penalizar a los participantes por las falsas alarmas, obligarles a decir «sí» un determinado porcentaje de veces, pedirles una valoración de la fiabilidad en lugar de limitarse a apuntar hacia arriba o hacia abajo con el pulgar, y hacer pruebas de respuesta múltiple en vez de test de verdadero o falso.

COSTES Y BENEFICIOS, Y ESTABLECIMIENTO DE UN CORTE

Con el trágico equilibrio entre aciertos y falsas alarmas (u omisiones y rechazos correctos), ¿qué ha de hacer un observador racional? Asumiendo por el momento que tenemos que limitarnos a los sentidos e instrumentos de medida de los que disponemos, junto con sus curvas de campana fastidiosamente superpuestas, la respuesta surge directamente de la teoría de la utilidad esperada (capítulo 6): depende de los beneficios de cada clase de conjetura correcta y los costes de cada clase de error.[7]

Regresemos al escenario en el que se originó la teoría de la detección de señales: el rastreo de la llegada de bombarderos a partir de los puntos del radar. Las cuatro posibilidades se presentan abajo. Cada fila representa un estado del mundo y cada columna una respuesta de nuestro operador de radar, con el resultado indicado en cada celda:

	«Sí»	**«No»**
Señal (bombardero)	Acierto (ciudad salvada)	Omisión (ciudad bombardeada)
Ruido (gaviotas)	Falsa alarma (misión desperdiciada, escalada de las tensiones)	Rechazo correcto (todo en calma)

Al decidir dónde fijar el criterio para responder, nuestro decisor tiene que ponderar los costes combinados (la utilidad esperada) de cada columna.[8] Las respuestas «sí» salvarán la ciudad objetivo de los bombardeos cuando esté siendo verdaderamente atacada (un acierto), lo cual supone un beneficio enorme, en tanto que se incurrirá en unos costes moderados cuando no lo esté (una falsa alarma), entre los que figuran el desperdicio de enviar aviones interceptores sin motivo alguno, junto con el temor de la población y las tensiones en el exterior. Las respuestas «no» expondrán a la ciudad al ataque cuando haya uno (una omisión), un coste masivo, mientras que preservarán la bendita paz y la tranquilidad cuando no haya ninguno (un rechazo correcto). El balance global parecería requerir un criterio de respuesta bajo, con una disposición a disparar con relativa

facilidad: el envío innecesario de interceptores unos cuantos días parecería un precio bajo a cambio del día en que evitarían que la ciudad fuese bombardeada.

El cálculo sería diferente si los costes fueran distintos. Supongamos que la respuesta no fuese el envío de aviones para interceptar a los bombarderos, sino el envío de misiles balísticos intercontinentales con ojivas nucleares para destruir las ciudades del enemigo, garantizando una Tercera Guerra Mundial termonuclear. En ese caso, el catastrófico coste de una falsa alarma exigiría la certeza absoluta de estar siendo atacado antes de responder, lo cual significa establecer un criterio de respuesta altísimo.

También son relevantes las tasas base de los bombarderos y las gaviotas que originan esos puntos (las probabilidades previas bayesianas). Si las gaviotas fueran comunes, pero los bombarderos raros, se precisaría un criterio elevado (no disparar con facilidad), y viceversa.

Como vimos en el capítulo anterior, nos enfrentamos al mismo dilema en una escala personal al decidir si someternos o no a cirugía en respuesta a un resultado ambiguo de una prueba del cáncer.

	«Sí»	**«No»**
Señal (cáncer)	Acierto (vida salvada)	Omisión (muerte)
Ruido (quiste benigno)	Falsa alarma (dolor, desfiguración, gastos)	Rechazo correcto (vida habitual)

Así las cosas, ¿dónde debería situar exactamente el criterio un decisor racional, esto es, un *observador ideal*, en la jerga de la teoría? La respuesta es: en el punto que maximizara la utilidad esperada del observador.[9] Este es fácil de calcular en el laboratorio, donde el experimentador controla el número de ensayos con un pitido (la señal) y un no pitido (el ruido), paga al participante por cada acierto y rechazo correcto, y le multa por cada omisión y falsa alarma. Entonces un participante hipotético que desease ganar el máximo dinero posible fijaría su criterio de acuerdo con esta fórmula, en la que los valores son las recompensas y las sanciones:

$$\beta = \frac{(\text{valor de un rechazo correcto} - \text{valor de una falsa alarma}) \times P(\text{ruido})}{(\text{Valor de un acierto} - \text{valor de una omisión}) \times P(\text{señal})}$$

El álgebra exacta es menos importante que la simple observación de lo que está arriba y abajo de la relación y lo que hay a cada lado del signo menos. Un observador ideal fijaría un criterio más elevado (necesitaría mejores pruebas antes de decir «sí») en la medida en que el ruido sea más probable que una señal (una baja probabilidad bayesiana *a priori*). Es una cuestión de sentido común: si las señales son raras, deberías decir «sí» con menos frecuencia. Asimismo, situaría un listón más alto cuando las recompensas por los aciertos son más bajas o por los rechazos correctos son más altas, mientras que las sanciones por las falsas alarmas son más altas o por las omisiones son más bajas. Una vez más, es de sentido común: si estás pagando grandes multas por las falsas alarmas, deberías ser más cauteloso a la hora de decir «sí», pero si estás obteniendo beneficios inesperados por los aciertos, deberías estar más dispuesto a hacerlo. En los experimentos de laboratorio, los participantes gravitan intuitivamente hacia lo óptimo.

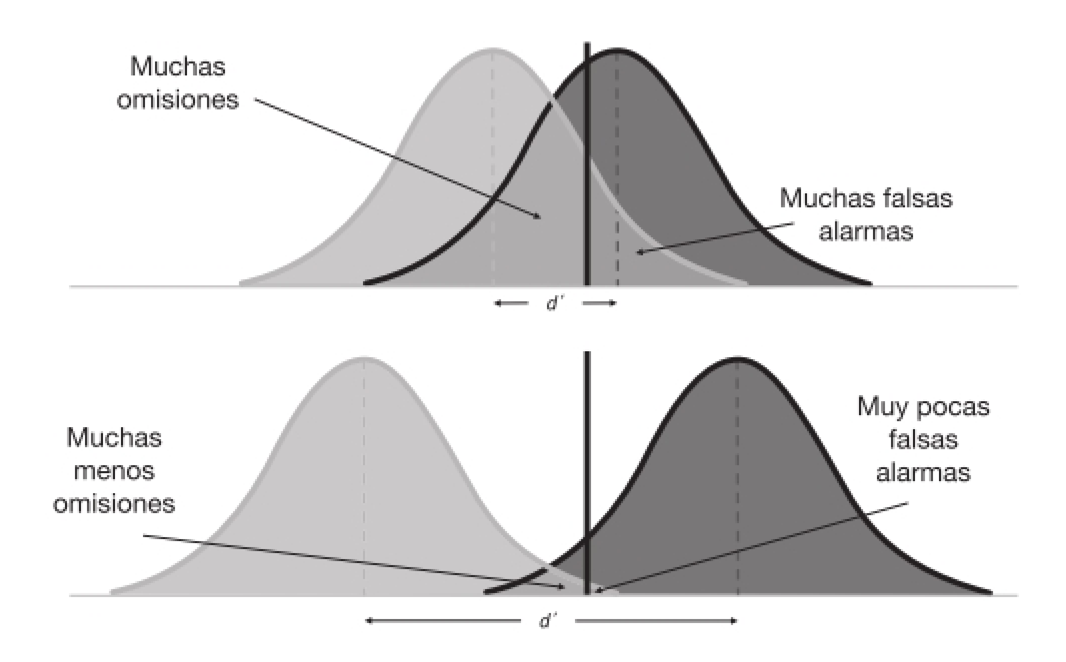
Cuando se trata de decisiones concernientes a la vida y la muerte, el dolor y la desfiguración, o la salvación o la destrucción de la civilización, la asignación de números a los costes resulta obviamente más problemática. No obstante, los dilemas siguen siendo igual de angustiosos si no les asignamos un valor, y la ponderación de las cuatro casillas, incluso con un sentido rudimentario de qué costes son monstruosos y cuáles soportables, puede tornar más consistentes y justificables las decisiones.

SENSIBILIDAD VERSUS SESGO DE RESPUESTA

Los equilibrios entre las omisiones y las falsas alarmas son angustiosos, y pueden instilar una visión trágica de la condición humana. ¿Estamos los mortales perpetuamente condenados a elegir entre el espantoso coste de la inacción equivocada (una ciudad bombardeada, un cáncer que se deja que se extienda) y el terrible coste de la acción equivocada (una provocación ruinosa, una cirugía desfigurante)? La teoría

de la detección de señales dice que lo estamos, pero nos muestra asimismo cómo mitigar la tragedia. Podemos alterar el equilibrio incrementando la *sensibilidad* de nuestras observaciones. Los costes de una tarea de detección de señales dependen de dos parámetros: dónde establecemos el corte (nuestro sesgo de respuesta, criterio, disposición a disparar o ß), y la distancia entre las distribuciones de las señales y del ruido, denominada *sensibilidad*, simbolizada como *d’*, que se pronuncia «d prima».[10]

Imaginemos que perfeccionásemos nuestro radar de forma que este eliminase las gaviotas o, en el peor de los casos, las registrase como nieve ligera, mientras que mostrase los bombarderos como grandes puntos brillantes. Eso significa que las curvas de campana para el ruido y la señal se apartarían (diagrama inferior). Esto implica, a su vez, que independientemente de dónde situásemos el corte de respuesta, tendríamos menos omisiones *y* menos falsas alarmas:



Y, por las leyes de la aritmética, lograríamos una proporción mayor de aciertos y rechazos correctos. Mientras que deslizar el corte hacia atrás y hacia delante supone intercambiar trágicamente un error por otro, separar las dos curvas (mediante mejores instrumentos, diagnósticos más sensibles, investigaciones forenses más fiables) supone un bien absoluto, reduciendo

errores de ambos tipos. El aumento de la sensibilidad debería ser siempre nuestra aspiración en los retos de detección de señales, lo cual nos conduce a una de sus aplicaciones más importantes.

DETECCIÓN DE SEÑALES EN LOS TRIBUNALES

La investigación de un delito es una tarea de detección de señales. Un juez, un jurado o un tribunal disciplinario se enfrentan a las evidencias sobre las posibles actividades ilícitas de un acusado. Las evidencias varían en fuerza, y un determinado conjunto de pruebas podría haber surgido del hecho de que el acusado hubiera cometido el delito (una señal) o de algún otro elemento, como que lo hubiese cometido otra persona o que no se hubiera producido delito alguno (ruido).

Las distribuciones de las pruebas se solapan más de lo que la mayoría de la gente aprecia. La llegada de la huella genética o prueba de ADN (un salto gigantesco en la sensibilidad) ha demostrado que muchas personas inocentes, algunas de ellas en el corredor de la muerte, habían sido condenadas sobre la base de evidencias que podrían haber provenido del ruido casi con tanta frecuencia como de una señal. Lo más notorio es el testimonio de los testigos oculares: las investigaciones de Elizabeth Loftus y otros psicólogos cognitivos han revelado que la gente recuerda habitualmente y con seguridad haber visto cosas que jamás ocurrieron.[11] Y la mayoría de los métodos aparentemente científicos y técnicos que aparecen en *CSI* y otros programas forenses de televisión nunca han sido validados adecuadamente, pero son promovidos por autoproclamados expertos con todo su exceso de confianza y sus sesgos de confirmación. Entre estos figuran los análisis de balística, las marcas de mordeduras, las fibras, los cabellos, las huellas de calzado o de neumáticos, las marcas de herramientas, la caligrafía, las salpicaduras de sangre, los aceleradores de incendios e incluso las huellas dactilares.[12] El ADN es la técnica forense más fiable, pero recordemos la diferencia entre una propensión y una frecuencia: un cierto porcentaje de testimonios de ADN está corrompido por muestras contaminadas, chapuzas en el etiquetado y otros errores humanos.

Un jurado que se enfrenta a pruebas ruidosas ha de aplicar un criterio y dictar un veredicto de «sí» o «no». Su matriz de decisión tiene costes y beneficios que se calculan en divisas prácticas y morales: los malhechores que son retirados de las calles o a quienes se permite aprovecharse de los demás, el valor abstracto de la justicia impuesto o malogrado.

	«Condenar»	**«Absolver»**
Señal (culpable)	Acierto (se hace justicia; criminal incapacitado)	Omisión (justicia denegada, criminal libre para aprovecharse de los demás)
Ruido (inocente)	Falsa alarma (error judicial; un inocente castigado)	Rechazo correcto (se hace justicia; aunque con costas judiciales)

Como vimos en la discusión de las tasas base prohibidas (capítulo 5), nadie toleraría un sistema judicial que funcionara puramente sobre la base práctica de los costes y los beneficios para la sociedad; insistimos en la justicia para el individuo. Ahora bien, dado que los jurados carecen de la omnisciencia divina, ¿cómo deberíamos compensar las injusticias inconmensurables de una condena falsa o de una absolución falsa? En el lenguaje de la detección de señales, ¿dónde situamos el criterio de respuesta?

La presunción habitual ha sido asignar un elevado coste moral a las falsas alarmas. Como dice el jurista William Blackstone (1723-1780) en su doctrina epónima: «Es preferible que se escapen diez personas culpables a que sufra un inocente». Y, de este modo, los jurados de los procesos penales aplican una «presunción de inocencia» y pueden condenar únicamente si el acusado es «culpable más allá de toda duda razonable» (un listón alto para ß, el criterio o sesgo de respuesta). No pueden condenar basándose en una mera «preponderancia de la prueba», también conocida como «el 50 % más una pluma».

La proporción de 10:1 de Blackstone es arbitraria, por supuesto, pero el desequilibrio es eminentemente defendible. En una democracia, la libertad es el valor por defecto, y la coerción gubernamental una onerosa excepción que ha de satisfacer una alta carga de justificación, habida cuenta

del impresionante poder del Estado y su constante tentación de tiranizar. El castigo al inocente, particularmente con la muerte, sacude la conciencia como no lo hace el hecho de no castigar al culpable. Un sistema que no apunta caprichosamente contra la gente para causarle la ruina marca la diferencia entre la justicia y el terror.

Como sucede con todos los ajustes de un criterio de respuesta, el ajuste basado en la proporción de Blackstone depende de la valoración de los cuatro resultados, que puede ser controvertida. A raíz del 11S, la Administración de George W. Bush creía que el catastrófico coste de un importante atentado terrorista justificaba el uso del *interrogatorio mejorado*, un eufemismo para la tortura, y tenía más peso que el coste moral de las confesiones falsas por parte de inocentes torturados.[13] En 2011, el Departamento de Educación de Estados Unidos desató una tormenta con una nueva directriz (rescindida desde entonces), en virtud de la cual las universidades han de condenar a los estudiantes acusados de conductas sexuales inapropiadas basándose en la preponderancia de la prueba.[14] Algunos defensores de esas medidas reconocían el riesgo que entrañaban, pero aducían que las infracciones sexuales son tan atroces que merecía la pena pagar el precio de condenar a unos cuantos inocentes.[15]

No existe ninguna respuesta «correcta» para estas preguntas de valoración moral, pero podemos utilizar la reflexión acerca de la detección de señales para determinar si nuestras prácticas son consistentes con nuestros valores. Supongamos que creemos que no más del 1 % de personas culpables deberían ser absueltas y no más del 1 % de las inocentes deberían ser condenadas. Supongamos asimismo que los jurados fuesen observadores ideales que aplicasen óptimamente la teoría de la detección de señales. ¿Cuán fuertes tendrían que ser las evidencias para cumplir esos objetivos? Para ser exactos, ¿cuán grande ha de ser *d’*, a saber, la distancia entre las distribuciones de la señal (culpable) y el ruido (inocente)? La distancia puede medirse en desviaciones típicas, la estimación más común de la variabilidad (visualmente, se corresponde con la anchura de la curva de campana, es decir, la distancia horizontal desde el punto medio hasta el punto de inflexión, donde lo convexo pasa a ser cóncavo).

Los psicólogos Hal Arkes y Barbara Mellers efectuaron los cálculos y concluyeron que, para cumplir esos objetivos, la *d’* de la fuerza de las evidencias ha de ser 4,7, lo cual significa que casi cinco desviaciones típicas separan las pruebas para los culpables de las pruebas para los inocentes.[16] Esto supone un nivel olímpico de sensibilidad que no se alcanza ni siquiera mediante nuestras tecnologías médicas más sofisticadas. Si estuviéramos dispuestos a relajar nuestros estándares y condenar hasta un 5 % de los inocentes y absolver un 5 % de los culpables, *d’* «solo» tendría que ser 3,3 desviaciones típicas, que sigue siendo un nivel de sensibilidad digno de la princesa y el guisante.

¿Quiere esto decir que nuestras aspiraciones morales a la justicia aventajan a nuestras facultades probatorias? Casi con certeza. Arkes y Mellers investigaron una muestra de estudiantes para ver cuáles son realmente esas aspiraciones. Los estudiantes aventuraron que una sociedad justa no debería condenar a más del 5 % de los inocentes ni absolver a más del 8 % de los culpables. Una muestra de jueces tenía intuiciones similares (no podemos decir si esta es una proporción más o menos estricta que la de Blackstone, porque no sabemos qué porcentaje de los acusados son realmente culpables). Esas aspiraciones exigen una *d’* equivalente a tres, es decir, que las pruebas dejadas por los acusados culpables tendrían que ser tres desviaciones típicas más fuertes que las pruebas dejadas por los inocentes.

¿Cuán realista es esto? Arkes y Mellers bucearon en la literatura sobre la sensibilidad de varios test y técnicas, y descubrieron que la respuesta es que no mucho. Cuando se pide a los sujetos que distingan a los mentirosos de quienes dicen la verdad, su *d’* es aproximadamente cero, lo cual quiere decir que son incapaces de hacerlo. El testimonio de los testigos presenciales es mejor que eso, pero no demasiado, situándose en un modesto 0,8. Los detectores de mentiras mecánicos, es decir, las pruebas de polígrafo, son mejores aún, en torno a 1,5, pero resultan inadmisibles en la mayoría de los tribunales.[17] Pasando de las pruebas forenses a otras clases de test para calibrar nuestras expectativas, hallaron un valor para *d’* de 0,7 aproximadamente en las pruebas de selección de personal militar; 0,8-1,7 para las predicciones meteorológicas; 1,3 para las mamografías; y 2,4-2,9

para los TAC de lesiones cerebrales (es cierto que estimadas con tecnologías de finales del siglo XX; todas deberían ser más altas en la actualidad).

Supongamos que la calidad típica de las pruebas en un juicio con jurado tiene una *d’* de uno (es decir, una desviación estándar más alta para los acusados culpables que para los inocentes). Si los jurados adoptan un criterio de respuesta estricto, anclado, pongamos por caso, por la creencia previa en que un tercio de los acusados son culpables, absolverán al 58 % de los acusados culpables y condenarán al 12 % de los inocentes. Si adoptan uno laxo, correspondiente a la creencia previa en que dos tercios de los acusados son culpables, absolverán al 12 % de los acusados culpables y condenarán al 58 % de los inocentes. La desoladora conclusión es que los jurados absuelven a muchas más personas culpables y condenan a muchas más inocentes de las que cualquiera de nosotros consideramos aceptables.

Ahora bien, el sistema penal puede hacer un trato mejor que ese con el diablo. La mayoría de los casos no llegan a juicio, sino que son desestimados porque las pruebas son muy débiles o se alcanza un acuerdo previo (idealmente) porque las evidencias son muy fuertes. Sin embargo, la mentalidad de la detección de señales podría encaminar nuestros debates sobre los procedimientos judiciales hacia una mayor justicia. Hoy en día, muchas de las campañas son ingenuas en lo que respecta al equilibrio entre aciertos y falsas alarmas, y tratan la posibilidad de las acusaciones falsas como algo inconcebible, como si los jurados fuesen infalibles. Muchos defensores de la justicia abogan por bajar el umbral de decisión. Meter a más criminales entre rejas. Creer a la mujer. Vigilar a los terroristas y encerrarlos antes de que ataquen. Si alguien quita una vida, merece perder la suya. Ahora bien, por necesidad matemática, rebajar el criterio de respuesta no puede sino sustituir una clase de injusticia por otra. Los argumentos podrían reformularse en estos términos: meter a más personas inocentes entre rejas; acusar de violación a más hombres libres de culpa; encerrar a jóvenes inofensivos que hablan más de la cuenta en los medios sociales; ejecutar a más inocentes.[18] Por sí mismas, estas paráfrasis no refutan los argumentos. En un momento dado, un sistema puede estar privilegiando de hecho al acusado por encima de sus posibles víctimas o

viceversa, y puede precisar un ajuste. Y si los humanos, que distan de ser omniscientes, han de tener un sistema judicial, deben afrontar la lúgubre necesidad de que algunos inocentes sean castigados.

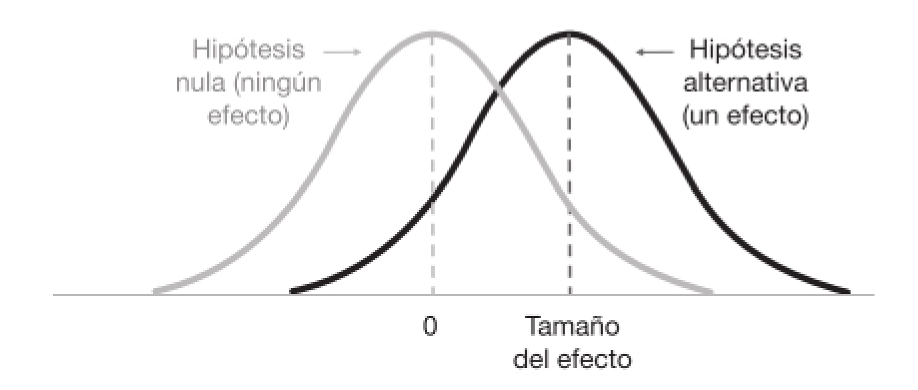
Con todo, la conciencia de los trágicos inconvenientes a la hora de distinguir las señales del ruido puede reportar mayor justicia. Nos obliga a afrontar la enormidad de los castigos severos como la pena de muerte y las largas condenas, que no son solo crueles para el culpable, sino que también afectarán inevitablemente a los inocentes. Y nos dice que la auténtica búsqueda de la justicia debería consistir en incrementar la sensibilidad del sistema, no su sesgo: buscar investigaciones forenses más precisas, protocolos más justos para los interrogatorios y los testimonios, restricciones al fanatismo acusador y otras salvaguardas contra los errores judiciales de ambas clases.

DETECCIÓN DE SEÑALES Y SIGNIFICACIÓN ESTADÍSTICA

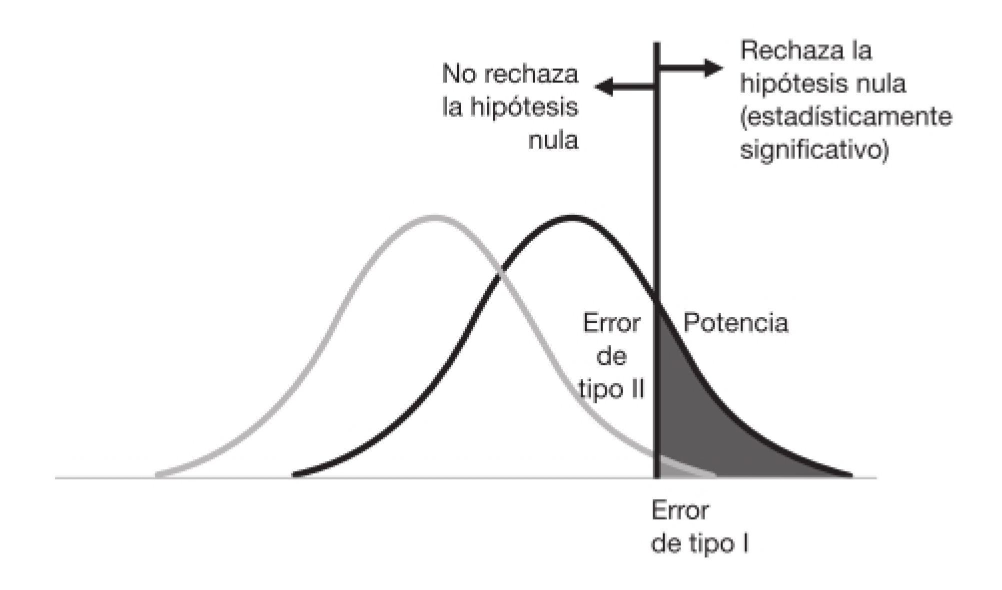
El equilibrio entre los aciertos y las falsas alarmas es inherente a cualquier decisión basada en evidencias imperfectas, lo cual significa que se cierne sobre todo juicio humano. Mencionaré uno más: las decisiones acerca de si un descubrimiento científico debería autorizar una conclusión sobre la verdad de una hipótesis. En este terreno, la teoría de la detección de señales aparece bajo el disfraz de la teoría estadística de la decisión.[19]

La mayoría de las personas científicamente informadas han oído hablar de la *significación estadística*, pues se menciona con frecuencia en las noticias sobre los descubrimientos en medicina, epidemiología y ciencias sociales. Tiene esencialmente la misma base matemática que la teoría de la detección de señales, iniciada por los estadísticos Jerzy Neyman (1894-1981) y Egon Pearson (1895-1980). Ver la conexión te ayudará a evitar un error garrafal que incluso una mayoría de científicos cometen de manera rutinaria. A todo estudiante de Estadística se le advierte que *significación estadística* es un concepto técnico que no debería confundirse con *significación* en el sentido ordinario de algo digno de atención o relevante. Pero la mayoría de la gente está mal informada sobre su auténtico significado.

Supongamos que un científico observa ciertas cosas del mundo y convierte sus mediciones en datos que representan el efecto en el que está interesado, como la diferencia en los síntomas entre el grupo que tomó el medicamento y el grupo que ingirió el placebo, o la diferencia en las destrezas verbales entre niños y niñas, o la mejora en los resultados de los exámenes de los estudiantes después de inscribirse en un programa de enriquecimiento. Si el número es cero, eso significa que no se produce efecto alguno; mayor que cero, un posible eureka. Pero, al ser como son las cobayas humanas, los datos son ruidosos, y una puntuación media por encima de cero puede significar que existe una diferencia real en el mundo, o podría tratarse de un error muestral, cuestión de suerte. Volvamos a la visión del ojo de Dios y representemos la distribución de resultados que obtendría el científico si no hubiera diferencia en la realidad —la llamada *hipótesis nula*—, y la distribución de resultados que obtendría si ocurriera algo, un efecto de un tamaño determinado. Las distribuciones se solapan: eso es lo que hace difícil la ciencia. El diagrama debería resultar familiar:



La hipótesis nula es el ruido; la hipótesis alternativa es la señal. El tamaño del efecto es como la sensibilidad, y determina cuán fácil es distinguir la señal del ruido. El científico necesita aplicar algún criterio o sesgo de respuesta antes de abrir el champán, llamado *valor crítico*: por debajo del valor crítico, no logra rechazar la hipótesis nula y ahoga sus penas; por encima del valor crítico, la rechaza y lo celebra, declara que el efecto es *estadísticamente significativo*.



Pero ¿dónde debería ubicarse el valor crítico? El científico debe compensar dos clases de errores. Podría rechazar la hipótesis nula cuando esta sea verdadera, esto es, una falsa alarma o, en el argot de la teoría estadística de la decisión, un *error de tipo I*. O podría no rechazar la hipótesis nula cuando esta sea falsa: una omisión o, en la jerga, un *error de tipo II*. Ambas cosas son malas: un error de tipo I introduce la falsedad en el registro científico; un error de tipo II representa un esfuerzo desperdiciado y un derroche de dinero. Esto sucede cuando la metodología no ha sido diseñada con suficiente «potencia» (índice de aciertos, o uno menos el índice de errores de tipo II) para detectar el efecto.

Ahora bien, en la noche de los tiempos se decidió —no está del todo claro quién lo hizo— que un error de tipo I (proclamar un efecto cuando no existe ninguno) resulta especialmente perjudicial para la empresa científica, que puede tolerar solamente un cierto número de ellos: el 5 % de los estudios en los que la hipótesis nula es verdadera, para ser exactos. Y de ese modo surgió la convención de que los científicos deberían adoptar un nivel crítico que garantice que la probabilidad de rechazar la hipótesis nula cuando esta sea verdadera es menor del 5 %: la codiciada p <0,05 (aunque cabría pensar que los costes de un error del tipo II también deberían tenerse en cuenta, como sucede en la teoría de la detección de señales, por alguna razón histórica igualmente oscura nunca lo fueron).

Eso es lo que quiere decir *significación estadística*: es una forma de mantener la proporción de falsas declaraciones de descubrimientos por debajo de un límite arbitrario. Entonces, si has obtenido un resultado estadísticamente significativo con p <0,05, eso quiere decir que puedes concluir lo siguiente, ¿verdad?

- La probabilidad de que la hipótesis nula sea verdadera es menor de 0,05.
- La probabilidad de que haya un efecto es mayor de 0,95.
- Si rechazas la hipótesis nula, hay una probabilidad menor de 0,05 de que tomes la decisión equivocada.
- Si replicas el estudio, la probabilidad de que tengas éxito es mayor de 0,95.

Esto es lo que piensan el 90 % de los profesores de Psicología, incluido el 80 % de los que enseñan Estadística.[20] Pero están total y absolutamente equivocados. Si has seguido la discusión de este capítulo y del capítulo 5, puedes entender por qué. La significación estadística es una *verosimilitud* bayesiana: la probabilidad de obtener los datos dada la hipótesis (en este caso, la hipótesis nula).[21] Pero cada uno de estos enunciados es una probabilidad *a posteriori* bayesiana: la probabilidad de la hipótesis dados los datos. Eso es lo que deseamos en última instancia —es la finalidad que persigue un estudio—, pero no es lo que arroja una prueba de significación. Si recuerdas por qué Irwin no tiene una enfermedad hepática, por qué las viviendas particulares no son necesariamente peligrosas y por qué el papa no es un extraterrestre, sabes que estas dos probabilidades condicionales no deben intercambiarse. El científico no puede utilizar una prueba de significación para evaluar si la hipótesis nula es verdadera o falsa, a menos que considere asimismo la probabilidad previa: su mejor estimación de la probabilidad de que la hipótesis nula sea verdadera antes de realizar el experimento. Y, en las matemáticas de las pruebas de significación de la hipótesis nula, no se encuentra en ninguna parte una probabilidad previa bayesiana.

La mayoría de los científicos sociales están tan enfrascados en el ritual de las pruebas de significación, desde tan temprano en sus carreras, que han olvidado su auténtica lógica. Yo me percaté de ello cuando colaboraba con una lingüista teórica, Jane Grimshaw, que estaba estudiando Estadística por su cuenta y me preguntó: «A ver si me aclaro. Lo único que demuestran estas pruebas es que, cuando algún efecto no existe, uno de cada veinte científicos que lo busquen afirmará falsamente que sí existe. ¿Cómo puedes estar tan seguro de que no serás *tú*?». La respuesta sincera es que no puedes. Su escepticismo anticipaba otra explicación del embrollo de la replicabilidad. Supongamos que, como los cazadores del Snark de Lewis Carroll, veinte científicos andan a la caza de una fantasía. Diecinueve archivan sus resultados nulos en un cajón y el suficientemente afortunado (o desafortunado) como para cometer el error del tipo I publica su «descubrimiento».[22] En una tira cómica de *XKCD*, un par de científicos comprueban si existe una correlación entre las gominolas y el acné separadamente para cada uno de veinte colores, y se hacen famosos por conectar las gominolas verdes con el acné con $p < 0{,}05$.[23] Los científicos han entendido por fin el chiste, se están acostumbrando a publicar resultados nulos y han desarrollado técnicas para compensar el sesgo de publicación cuando revisan la literatura en un metaanálisis, un estudio de estudios. Los resultados nulos brillan por su ausencia, y el analista puede detectar tanto la nada ausente como la nada presente.[24]

El escandaloso malentendido de las pruebas de significación denota un anhelo humano. Los filósofos desde Hume han advertido que la inducción —extraer una generalización a partir de las observaciones— es una clase de inferencia intrínsecamente incierta.[25] Cabe trazar un número infinito de curvas que pasen por un conjunto finito de puntos; un número ilimitado de teorías son lógicamente consistentes con cualquier cuerpo de datos. Las herramientas de la racionalidad explicadas en estos capítulos ofrecen diferentes formas de lidiar con esta desgracia cósmica. La teoría estadística de la decisión no puede determinar la verdad, pero puede limitar el daño de las dos clases de error. El razonamiento bayesiano puede ajustar nuestra creencia en la verdad, pero ha de comenzar con una probabilidad previa,

con todo el juicio subjetivo que esta comporta. Ninguno de ambos proporciona lo que todo el mundo desea: un algoritmo preparado para determinar la verdad.

CAPÍTULO 8

El yo y los otros
(Teoría de juegos)

Tu maíz está maduro hoy; el mío lo estará mañana. Es conveniente para ambos que yo trabaje contigo hoy y que tú me ayudes mañana. No siento afecto por ti y sé que tú tampoco lo sientes por mí. Por tanto, no me esforzaré en tu beneficio y, si trabajara contigo en mi beneficio, sé que me decepcionarías y que no podría depender de tu gratitud. Así sucede que te dejo solo en tu trabajo y tú me tratas de la misma manera. Las estaciones cambian y ambos perdemos nuestra cosecha por falta de confianza y seguridad mutuas.

DAVID HUME[1]

No hace mucho tiempo tuve una discusión amistosa con un colega acerca de los mensajes que debería enviar nuestra universidad sobre el cambio climático. El profesor J. argüía que solo tenemos que persuadir a la gente de que la reducción de sus emisiones de gases de efecto invernadero redundará en su propio interés, ya que un planeta más caliente traería consigo inundaciones, huracanes, incendios forestales y otros desastres que empeorarían sus vidas. Yo respondía que *no* era en interés propio, pues ningún sacrificio individual puede evitar por sí solo el cambio climático, en tanto que el sacrificador sudaría en verano, tiritaría en invierno y esperaría el autobús bajo la lluvia mientras sus vecinos contaminantes estarían cómodos y secos. Solamente si *todos* eliminasen sus emisiones se beneficiaría *cualquiera*, y la única forma de que a cualquiera le interesase hacerlo era que las energías limpias fuesen más baratas para todos (mediante los avances tecnológicos) y las energías sucias fuesen más caras (mediante la tarificación del carbono). Mi colega tenía razón: en cierto sentido, es irracional arruinar el planeta. Pero no fui capaz de convencer al doctor J. de que, en otro sentido, es, trágicamente, muy racional.

En ese momento me percaté de que en la cosmovisión del buen doctor estaba ausente un concepto crucial: la teoría de juegos, el análisis de cómo tomar decisiones racionales cuando los beneficios dependen de las decisiones racionales de los *otros*.

La teoría de juegos fue presentada al mundo por Von Neumann y Morgenstern en el mismo libro en el que explicaban la utilidad esperada y la elección racional.[2] Pero, a diferencia de los dilemas en los que nos arriesgamos contra una descerebrada rueda de la fortuna y las mejores estrategias resultan muy intuitivas, la teoría de juegos se ocupa de dilemas en los que nos enfrentamos a decisores igualmente astutos, y los resultados pueden poner patas arriba nuestras intuiciones. Los juegos de la vida a veces no dejan a los actores racionales más opción que hacer cosas que empeoren tanto su situación como la de todos los demás; actuar al azar, de manera arbitraria o fuera de control; cultivar simpatías y alimentar agravios; someterse voluntariamente a sanciones y castigos; y, a veces, negarse

rotundamente a jugar. La teoría de juegos desvela la extraña racionalidad que subyace a muchas de las perversidades de la vida social y política, y, como veremos en un capítulo posterior, contribuye a explicar el misterio central de este libro: cómo una especie racional puede ser tan irracional.

UN JUEGO DE SUMA CERO: PIEDRA, PAPEL O TIJERA

El dilema por antonomasia de la teoría de juegos, que pone de manifiesto cómo el beneficio de una elección depende de la elección del otro, es el juego de piedra, papel o tijera.[3] Dos jugadores muestran simultáneamente un gesto con la mano (dos dedos para tijera, la palma plana para papel, el puño cerrado para piedra) y el ganador se decide mediante la regla: «La tijera corta el papel, el papel cubre la piedra, la piedra despunta la tijera». El juego puede presentarse como una matriz en la que las elecciones posibles de la primera jugadora, Amanda, se muestran como filas, las elecciones del segundo, Brad, se muestran como columnas, y los resultados se indican en cada celda, los de Amanda en el ángulo inferior izquierdo y los de Brad en el superior derecho. Asignemos valores numéricos a los resultados: 1 para una victoria, –1 para una derrota, 0 para un empate.

Elecciones de Amanda \ Elecciones de Brad	Tijera	Papel	Piedra
Tijera	Brad: Empata 0 Amanda: Empata 0	Brad: Pierde –1 Amanda: Gana 1	Brad: Gana 1 Amanda: Pierde –1
Papel	Brad: Gana 1 Amanda: Pierde –1	Brad: Empata 0 Amanda: Empata 0	Brad: Pierde -1 Amanda: Gana 1
Piedra	Brad: Pierde –1 Amanda: Gana 1	Brad: Gana 1 Amanda: Pierde –1	Brad: Empata 0 Amanda: Empata 0

Los resultados de Amanda y los resultados de Brad suman cero en todas las celdas, lo cual nos brinda un término técnico que ha pasado de la teoría de juegos a la vida cotidiana: el juego de suma cero. La victoria de Amanda supone la derrota de Brad y viceversa. Están atrapados en un estado de puro conflicto, luchando por una tarta fija.

¿Qué jugada (fila) debería elegir Amanda? La técnica crucial en la teoría de juegos (y, de hecho, en la vida) consiste en ver el mundo desde el punto de vista del otro jugador. Amanda debe examinar las elecciones de Brad, las columnas, de una en una. Yendo de izquierda a derecha, si Brad elige tijera, ella debería elegir piedra. Si elige papel, ella debería elegir tijera. Y si elige piedra, ella debería elegir papel. No existe ninguna opción «dominante», que sea superior independientemente de lo que Brad haga y, por supuesto, ella no sabe lo que hará Brad.

Pero esto no significa que Amanda deba elegir una jugada arbitraria, por ejemplo, papel, y aferrarse a ella. Si lo hiciera, Brad lo captaría, escogería tijera y la vencería siempre. De hecho, incluso si ella se inclinara un poco hacia papel, eligiéndolo, pongamos por caso, el 40 % de las veces y las otras dos estrategias el 30 % cada una, Brad podría jugar tijera y vencerla cuatro veces de siete. La mejor estrategia de Amanda consiste en convertirse en una ruleta humana y elegir cada jugada al azar con la misma probabilidad, reprimiendo cualquier sesgo, inclinación o alejamiento de una división perfecta 1/31/3-1/3.

Dado que la tabla es simétrica siguiendo la diagonal, las maquinaciones de Brad son idénticas. Cuando este considera lo que Amanda podría hacer, fila a fila, no tiene motivo alguno para escoger una de sus jugadas por encima de las otras dos, y llegará a la misma estrategia «mixta», jugando cada opción con una probabilidad de 1/3. Si Brad se desviara de su estrategia, Amanda modificaría la suya para aprovecharlo y viceversa. Ambos están atrapados en el *equilibrio de Nash*, llamado así por el matemático John Nash, el protagonista de la película *A Beautiful Mind (Una mente maravillosa)*. Cada uno está jugando la mejor estrategia dada la mejor estrategia del rival; cualquier cambio unilateral les perjudicaría.

El descubrimiento de que, en determinadas situaciones, un agente racional ha de ser sobrehumanamente aleatorio es solo una de las conclusiones de la teoría de juegos que se nos antoja descabellada hasta que nos percatamos de que las situaciones no son infrecuentes en la vida. El equilibrio en piedra, papel o tijera se denomina enfrentamiento anticipatorio y los ejemplos son frecuentes en deportes como el tenis, el béisbol, el jóquey y el fútbol. El lanzador de un penalti en el fútbol puede lanzar a la derecha o a la izquierda, y el portero puede proteger la derecha o la izquierda; la impredecibilidad es una virtud cardinal. Los faroles en el póquer y los ataques sorpresa en la estrategia militar son también enfrentamientos anticipatorios. Incluso cuando una jugada no se elige literalmente al azar (presumiblemente, en 1944, los aliados no tiraron un dado antes de decidir si invadir Normandía o Calais), el jugador debe poner cara de póquer y reprimir cualquier gesto revelador o filtración, haciendo que la elección *parezca* aleatoria a sus oponentes. Los filósofos Liam Clegg y Dan Dennett han sostenido que la conducta humana es intrínsecamente impredecible no solo a causa del ruido neuronal aleatorio en el cerebro, sino también como una adaptación que torna más difícil que nuestros rivales se anticipen a nosotros.[4]

UN JUEGO DE SUMA NO CERO: EL DILEMA DEL VOLUNTARIO

Los actores racionales pueden acabar en enfrentamientos anticipatorios no solo en juegos que los enfrentan en una competición de suma cero, sino también en otros que los alinean con intereses comunes. Un ejemplo es el dilema del voluntario, que puede ilustrarse mediante el cuento medieval sobre poner el cascabel al gato. Un ratón propone a sus compañeros de casa que uno de ellos ponga un cascabel alrededor del cuello del gato mientras este duerme, con el fin de que los alerte cuando se acerque. El problema, por supuesto, es quién pondrá el cascabel al gato, arriesgándose a despertarlo y ser devorado. Entre los dilemas paralelos para los humanos figuran qué pasajero reducirá al secuestrador de un avión, qué espectador rescatará a una persona en apuros y qué oficinista rellenará la cafetera en una cocina comunitaria.[5] Todos quieren que alguien ayude, pero prefieren

no ser ellos. Si traducimos los beneficios y los costes a unidades numéricas, siendo cero lo peor que puede suceder, obtenemos la matriz que hay a continuación (técnicamente, debería ser un hipercubo con tantas dimensiones como jugadores, pero yo he fundido a todos menos al *yo* en una sola capa).

		Elecciones propias	
		Ayudar	Escaquearse
Elecciones ajenas	Ayudar	50 50	100 50
	Escaquearse	50 100	0 0

Una vez más, no existe ninguna estrategia dominante que haga fácil la elección: si un ratón supiera que los otros se escaquearían, entonces ayudaría, y viceversa. Pero si cada ratón decidiera si poner o no el cascabel al gato con una cierta probabilidad (que equiparase los beneficios esperados de poner el cascabel y escaquearse de los *otros* ratones), entonces los ratones caerían en un enfrentamiento anticipatorio, en el que cada uno estaría dispuesto a poner el cascabel, al tiempo que esperaría que otro ratón se le adelantara.

A diferencia de piedra, papel o tijera, el dilema del voluntario no es de suma cero: unos resultados son mejores para todos que otros. (Se trata de una situación en la que todos ganan, *win-win*, otro concepto de la teoría de juegos que se ha incorporado al lenguaje cotidiano.) Los ratones salen colectivamente perjudicados si ningún ratón se presenta voluntario y beneficiados si uno se ofrece, lo cual no garantiza que ese sea un final feliz, pues no hay ningún ratón jefe para enviar a uno de ellos al posible martirio por el bien del grupo. Antes bien, cada ratón tira el dado porque a ninguno le iría mejor cambiando unilateralmente de estrategia. Una vez más se encuentran en un equilibrio de Nash, un enfrentamiento en el que todos los jugadores se aferran a su mejor opción en respuesta a las mejores opciones de los demás.

LA CITA Y OTROS JUEGOS DE COORDINACIÓN

Un concurso de sálvese quien pueda, como piedra, papel o tijera, y un enfrentamiento inquietante e hipócrita, como el dilema del voluntario, implican cierto grado de competición. Sin embargo, en algunos juegos de la vida, todos ganan si consiguen averiguar cómo hacerlo. Estos se denominan juegos de coordinación, como la cita. Caitlin y Dan disfrutan de su mutua compañía y planean tomar café una tarde, pero el teléfono de Caitlin se apaga antes de que puedan decidirse por Starbucks o Peet’s. Cada uno tiene una ligera preferencia, pero ambos preferirían reunirse en cualquier lugar antes que privarse de la cita. La matriz tiene dos equilibrios, las celdas superior izquierda e inferior derecha, correspondientes a su coordinación en la misma elección (técnicamente, sus distintas preferencias introducen una pizca de competición en el escenario, pero podemos ignorarla por ahora).

		Elecciones de Dan	
		Peet’s	Starbucks
Elecciones de Caitlin	Peet’s	100 95	0 0
	Starbucks	0 0	95 100

Caitlin sabe que Dan prefiere Peet’s y decide presentarse allí, pero Dan sabe que Caitlin prefiere Starbucks, así que planea acudir *allí*. Caitlin, poniéndose en el lugar de Dan, anticipa la empatía de este, por lo que cambia su plan a Starbucks, y Dan, igualmente empático acerca de la empatía de Caitlin, cambia a Peet’s, hasta que advierte que ella ha anticipado su anticipación, y vuelve a cambiar a Starbucks. Y así sucesivamente, *ad infinitum*, sin que ninguno de los dos tenga un motivo para decidirse por algo que ambos desean.

Lo que necesitan es un *conocimiento común*, que en la teoría de juegos es un término técnico que designa algo que cada uno sabe que el otro sabe que él sabe, *ad infinitum*.[6] Aunque pueda parecer que el conocimiento común haría explotar la cabeza de alguien, uno no necesita meterse en el cráneo una serie infinita de «yo sé que ella sabe que yo sé que ella sabe...». Le basta con tener un sentido de que el conocimiento es «autoevidente» o está «ahí fuera» o «consta en acta». Esa intuición puede ser generada por una señal evidente que cada uno percibe con la conciencia del otro, como un diálogo en directo entre ellos. En muchos juegos, una mera promesa es «charla barata» y descartable (en un dilema del voluntario, por ejemplo, si un ratón anuncia que se niega a ofrecerse voluntario, confiando en que presionará para que lo haga otro, los demás ratones pueden descubrir su farol y escaquearse, sabiendo que este podría llenar el vacío). Sin embargo, en un juego de coordinación, a ambas partes les interesa terminar en el mismo lugar, por lo que una declaración de intenciones resulta creíble.

En ausencia de comunicación directa (como cuando un móvil se apaga), las partes pueden converger en su lugar en un *punto focal*: una elección mutuamente destacable, que cada parte se figura que la otra debe de haber notado y ser consciente de que ella también la ha advertido.[7] Si el Peet's estuviera cerca, o si hubiera aparecido recientemente en la conversación, o si fuera un hito familiar en la ciudad, eso podría bastar para que Caitlin y Dan salieran del atolladero, con independencia de qué sitio presumiese de mejor café con leche o de asientos más lujosos. En los juegos de coordinación, algo arbitrario, superficial e insignificante que atraiga la atención puede brindar la solución racional a un problema difícil de resolver.

Muchas de nuestras convenciones y estándares son soluciones para juegos de coordinación, sin más motivo para recomendarlas que el hecho de que todo el mundo se ha decantado por las mismas.[8] Conducir por la derecha, no trabajar los domingos, aceptar papel moneda, adoptar estándares tecnológicos (doscientos treinta voltios, Microsoft Word, el teclado QWERTY) son equilibrios en juegos de coordinación. Puede haber beneficios más altos con otros equilibrios, pero permanecemos encerrados

en los que tenemos porque no podemos llegar allí desde aquí. A menos que todo el mundo esté de acuerdo en cambiar a la vez, las sanciones por la descoordinación son demasiado elevadas.

Los puntos focales arbitrarios pueden figurar en las negociaciones. Una vez que un comprador y un vendedor han convergido en un rango de precios que hacen el trato más atractivo para ambos que marcharse, se hallan en una especie de juego de coordinación. Cualquiera de los dos equilibrios (sus respectivas ofertas actuales) resulta más atractivo que la falta absoluta de coordinación, pero cada uno de ellos le parece más atractivo a uno de los dos. Conforme cada parte cambia las recompensas, confiando en atraer a la otra hacia la celda de coordinación que le resulte más ventajosa, pueden buscar un punto focal que, aunque arbitrario, les ofrezca algo con lo que ambas estén de acuerdo, como una cifra redonda o una oferta que divida la diferencia en partes iguales. Como dice Thomas Schelling, que fue el primero en identificar los puntos focales en los juegos de coordinación: «El vendedor que calcula el precio mínimo de un automóvil en 35.017,63 dólares está suplicando prácticamente que le libren de los 17,63 dólares».[9] Análogamente, «si alguien ha estado exigiendo el 60 % y baja al 50 %, puede plantarse; si baja al 49 %, el otro asumirá que continuará deslizándose cuesta abajo».[10]

EL «GALLINA» Y LOS JUEGOS DE ESCALADA

Aunque la negociación posee elementos de un juego de coordinación, la capacidad de cada parte de amenazar a la otra levantándose de la mesa y perjudicando a ambas hace que se solape con otro célebre juego, el del gallina, que conocimos en el capítulo 2.[11] He aquí la matriz (como siempre, los números exactos son arbitrarios; solamente son significativas las diferencias).

		Elecciones de Buzz	
		Virar	Seguir recto
Elecciones de Jim	Virar	Anticlímax 0 Anticlímax 0	Victoria 1 «Gallina» −1
	Seguir recto	«Gallina» −1 Victoria 1	Choque −100 Choque −100

Los nombres de los jugadores proceden de *Rebel Without a Cause (Rebelde sin causa)*, pero el juego del gallina no es solo un pasatiempo suicida para adolescentes. Jugamos a este juego cuando conducimos o andamos por un camino estrecho y alguien se acerca de frente, y es preciso que uno de los dos ceda el paso, o cuando entablamos una negociación formal o informal. Entre los ejemplos públicos se incluyen la ejecución o el impago de una deuda, y las políticas arriesgadas en las relaciones internacionales, como la crisis de los misiles cubanos de 1962. El juego del gallina tiene un equilibrio de Nash en el que cada jugador se arriesga a mantenerse firme o, de lo contrario, vira, aunque en la vida real esta solución puede ser discutible, porque pueden enriquecerse las reglas del juego para incluir señalización y alteraciones en la estrategia fijada. En el capítulo 2 vimos cómo una ventaja paradójica puede favorecer a un jugador que esté visiblemente loco o fuera de control, tornando suficientemente creíbles sus amenazas como para obligar a ceder a su oponente, aunque con la sombra de la destrucción mutua cerniéndose sobre ellos si ambos enloquecen o pierden el control simultáneamente.[12]

Algunos juegos no consisten en un único encuentro en el que los jugadores hacen una sola jugada simultáneamente y luego muestran sus manos, sino en una serie de jugadas en las que cada uno responde al otro y las cuentas se ajustan al final. Uno de estos juegos secuenciales tiene implicaciones morbosas sorprendentes. Un juego de escalada puede ser ilustrado con una «subasta de dólares» en un eBay infernal.[13] Imaginemos una subasta con la diabólica regla de que el perdedor, no solo el ganador, tiene que pagar su última oferta. Supongamos que el artículo que se está

subastando es una baratija que se puede revender por un dólar. Amanda ofrece 5 centavos, esperando conseguir 95 centavos de beneficio. Pero, por supuesto, interviene Brad y ofrece 10 centavos, y así sucesivamente, hasta que la oferta de Amanda haya alcanzado los 95 centavos, lo cual recortaría su margen a 5 centavos, todavía rentables. A esas alturas puede parecer absurdo que Brad ofrezca un dólar para ganar un dólar, pero cubrir los gastos sería preferible a perder 90 centavos, que la perversa regla de la subasta le obligaría a pagar si se retirase. Más perversamente aún, Amanda se enfrenta ahora a la disyuntiva de perder 95 centavos si se rinde o perder 5 centavos si sube la puja, así que ofrece 1,05 dólares, que Brad, prefiriendo perder 10 centavos a un dólar, mejora con 1,10 dólares, y así sucesivamente. Quedan atrapados en una frenética espiral de pujar más alto que el otro, tirando cada vez más dinero hasta que uno de ellos quiebra y el otro disfruta la victoria pírrica de perder un poco menos.

La estrategia racional en medio de un juego de escalada consiste en reducir tus pérdidas y retirarte con una cierta probabilidad en cada jugada, confiando en que el otro postor, siendo igualmente racional, quiera rendirse antes. Queda plasmada en el dicho «no tires el dinero bueno después del malo» y en la primera ley de los agujeros: «Cuando estés en uno, deja de cavar». Una de las irracionalidades humanas citadas con más frecuencia es la falacia del costo hundido o irrecuperable, según la cual los individuos continúan invirtiendo en una empresa perdedora debido a lo que han invertido hasta el momento, en lugar de en previsión de lo que ganarán si siguen adelante. Conservar unas acciones que se desploman, aguantar una película aburrida, terminar una novela tediosa y permanecer en un mal matrimonio son ejemplos familiares. Es posible que la gente caiga presa de la falacia del costo hundido como un efecto colateral de jugar a la escalada (y al gallina), donde una reputación de mantenerse firme, por muy costoso que sea, podría convencer al otro jugador de que se retire primero.

El juego de escalada no es una exótica prueba de ingenio. La vida real nos plantea dilemas en los que sería aplicable el dicho «de perdidos, al río». Entre ellos figuran las huelgas laborales prolongadas, los procesos judiciales planteados como duelos y las guerras de desgaste en sentido literal, en las que cada nación introduce hombres y equipamiento militar en

las fauces de la maquinaria bélica, confiando en que el otro bando se agote primero.[14] La justificación habitual es «luchamos para que nuestros muchachos no hayan muerto en vano», un ejemplo de manual de la falacia del costo irrecuperable, pero también una táctica en la patética búsqueda de una victoria pírrica. Muchas de las guerras más sangrientas de la historia fueron guerras de desgaste, lo cual demuestra una vez más que la exasperante lógica de la teoría de juegos puede explicar algunas de las tragedias de la condición humana.[15] Aunque, una vez que uno está atrapado en un juego de escalada, persistir en cierta probabilidad puede ser la opción menos mala, la estrategia verdaderamente racional consiste de entrada en no jugar.

Esto incluye los juegos en los que puede que ni siquiera nos demos cuenta de que estamos jugando. Para mucha gente, uno de los beneficios de ganar una subasta es el mero placer de ganar. Dado que la emoción de la victoria y la agonía de la derrota son independientes de la cantidad de la puja ganadora y del valor del artículo, esto puede convertir cualquier subasta en un juego de escalada. Los subastadores explotan esta psicología creando suspense y colmando de elogios al ganador, y los usuarios de eBay aconsejan a los postores que decidan de antemano cuánto vale para ellos el artículo y no pujen más alto. Algunos venden una modalidad de autocontrol odiseico: pujan hasta un límite fijado de antemano por el postor, atándole al mástil por su propio bien durante el frenesí de un juego de escalada del ego.

EL DILEMA DEL PRISIONERO Y LA TRAGEDIA DE LOS COMUNES

Consideremos una trama familiar de *Law and Order (Ley y orden)*. Una fiscal detiene a dos cómplices de un delito en celdas separadas, carece de pruebas para condenarlos y les ofrece un trato. Si uno accede a testificar en contra del otro, quedará en libertad y su cómplice pasará diez años en la cárcel. Si cada uno delata al otro, ambos son condenados a seis años. Si mantienen su lealtad mutua y guardan silencio, solo puede condenarlos por un cargo menor y cumplirán seis meses.

Los resultados se muestran en la página siguiente. En las discusiones del dilema del prisionero, cooperar significa permanecer fiel al cómplice (no cooperar con la fiscal) y traicionar significa delatarle. Los resultados también tienen etiquetas mnemónicas y su orden de «maldad» es lo que define el dilema. Para cada jugador, el mejor resultado es traicionar mientras el otro coopera (la tentación), el peor es ser la víctima de semejante traición (el pago del bobo), la segunda peor opción es ser parte en una traición mutua (el castigo) y la segunda mejor es mantener la lealtad cuando el otro lo hace (la recompensa). El mejor y el peor resultado para la pareja considerada conjuntamente caen a lo largo de la otra diagonal: lo peor que puede sucederles colectivamente es la delación mutua y lo mejor la cooperación mutua.

		Brutus	
		Cooperar (silencio)	Traicionar (delación)
Lefty	Cooperar (silencio)	6 meses (recompensa) 6 meses (recompensa)	Libertad (tentación) 10 años (pago del bobo)
	Traicionar (delación)	10 años (pago del bobo) Libertad (tentación)	6 años (castigo) 6 años (castigo)

Considerando toda la tabla desde nuestra perspectiva omnipresente, resulta evidente dónde deberían tratar de terminar los cómplices. Ninguno de ambos puede contar con que el otro cargue con la culpa, por lo que el único objetivo sensato es la recompensa de la cooperación mutua. Lamentablemente para ellos, desde sus perspectivas terrenales, no se puede captar la tabla entera, porque la elección de su cómplice está fuera de su control. Lefty contempla las dos jugadas a su derecha y Brutus las dos que tiene debajo. Lefty tiene que planteárselo en estos términos: «Supongamos que él guarda silencio (coopera). Entonces me caerían seis meses si yo también guardara silencio y quedaría en libertad si cantara (traicionara).

Saldría ganando con la traición. Supongamos ahora que él me delata (traiciona). Entonces me caerían diez años si yo guardara silencio, pero solo seis si le delatara *a él*. En definitiva, esto significa que, si él coopera, a mí me compensa traicionar y, si él traiciona, a mí me compensa traicionar. Es pan comido». Mientras tanto, el bocadillo de pensamiento encima de la cabeza de Brutus contiene el mismo soliloquio. Ambos traicionan y son condenados a seis años en lugar de seis meses: el fruto amargo de la actuación racional de cada uno guiada por su propio interés. Ninguno de ellos tenía otra opción: se trata de un equilibrio de Nash. La traición es una estrategia dominante para ambos, que beneficia a cada uno de ellos independientemente de lo que haga el otro. Si uno de los dos hubiera sido sabio, o moral, o confiado o con visión de futuro, habría estado a merced del miedo y la tentación del otro. Incluso si su cómplice le hubiera asegurado que haría lo correcto, podría haberse tratado de palabrería barata, que no valiese ni el papel en el que estuviera escrita.

El dilema del prisionero es una tragedia común. Un marido y una mujer en proceso de divorcio contratan a abogados sin escrúpulos, pues cada uno de ellos teme que el otro le deje sin blanca, mientras las horas facturables van agotando los bienes conyugales. Las naciones enemigas hacen quebrar sus presupuestos respectivos en una carrera armamentista, con lo que ambas se empobrecen sin ser más seguras. Los ciclistas se dopan y corrompen el deporte porque de lo contrario les harían morder el polvo los rivales que, a su vez, se dopan.[16] Todos abarrotan una cinta transportadora de equipaje o se ponen en pie en un concierto de *rock*, estirándose para conseguir una vista mejor, y nadie acaba teniendo una vista mejor.

El dilema del prisionero no tiene solución, pero pueden cambiarse las reglas del juego. Una forma de hacerlo es que los jugadores lleguen a acuerdos de obligado cumplimiento antes de jugar, o se sometan al mandato de una autoridad, que modifique los resultados añadiendo una recompensa por la cooperación o una sanción por la traición. Supongamos que los cómplices prestan juramento de *omertà*, impuesto por el Padrino, de suerte que, si respetan un código de silencio, serán ascendidos a capos, mientras que, si lo rompen, acabarán durmiendo con los peces. Eso altera la matriz

de resultados, convirtiéndola en un juego diferente cuyo equilibrio es la cooperación mutua. A los cómplices les interesa prestar el juramento de antemano, aun cuando ello excluya su libertad de traicionar. Los actores racionales pueden escapar de un dilema del prisionero sometiéndose a contratos vinculantes y al imperio de la ley.

Otra variación del juego consiste en jugar repetidamente, recordando lo que hizo el cómplice en las rondas anteriores. En este caso, una pareja puede encontrar el camino a la bendita celda de cooperarcooperar y permanecer en ella jugando con una estrategia llamada toma y daca. Esta exige cooperar en la primera jugada y a partir de entonces pagar al cómplice con la misma moneda: cooperar si el compañero coopera, traicionar si él traiciona (en algunas versiones, concediéndole una traición gratis antes de traicionar, por si acaso se tratase de un desliz).

Los biólogos evolutivos han advertido que los animales sociales se encuentran con frecuencia en dilemas del prisionero iterados.[17] Un ejemplo es la mutua recompensa del acicalamiento recíproco, con la tentación de ser acicalado sin acicalar a su vez. Robert Trivers sugiere que el *Homo sapiens* ha desarrollado una serie de emociones morales que implementan el toma y daca, y nos permiten disfrutar de los beneficios de la cooperación.[18] La compasión nos impulsa a cooperar en la primera jugada; la gratitud, a pagar la cooperación con cooperación; la ira, a castigar la traición con traición; la culpa, a expiar nuestra traición antes de que esta sea castigada; el perdón, a impedir que la traición puntual de un compañero nos condene para siempre a la traición mutua. Muchos de los dramas de la vida social humana —las sagas de la compasión, la confianza, el favor, la deuda, la venganza, la gratitud, la culpa, la vergüenza, la traición, el chismorreo, la reputación— pueden interpretarse como un juego de estrategias en un dilema del prisionero iterado.[19] La cita que abre este capítulo muestra que Hume, una vez más, llegó primero a esa conclusión.

MUCHOS DE LOS DRAMAS de la vida política y económica pueden explicarse como dilemas del prisionero con más de dos jugadores, conocidos como juegos de bienes públicos.[20] Todos los miembros de una comunidad se benefician de los bienes públicos, como la iluminación, las

carreteras, el alcantarillado, la policía y las escuelas. Pero se benefician más aún si todos pagan por ellos, excepto ellos: una vez construida una farola, cualquiera puede verla. En una conmovedora versión ecológica llamada *tragedia de los comunes*, todos los pastores tienen un incentivo para añadir una oveja más a su rebaño y apacentarla en los terrenos comunales, pero, cuando todos ceban sus rebaños respectivos, la hierba se pasta más deprisa de lo que puede volver a crecer, y todas las ovejas mueren de hambre. El tráfico y la polución funcionan de la misma manera: mi decisión de conducir no atascará las carreteras ni ensuciará el aire, del mismo modo que mi decisión de coger el autobús no evitará estas cosas, pero, cuando todos conducen, todos sufren. Evadir impuestos, escatimar en una colecta, exprimir un recurso hasta agotarlo y resistirse a las medidas de salud pública —como la distancia social y llevar mascarilla durante una pandemia — son otros tantos ejemplos de traicionar en un juego de bienes públicos: ofrecen una tentación a aquellos que se lo permiten, un pago del bobo a quienes colaboran y conservan, y un castigo común cuando todos traicionan.

Por regresar al ejemplo con el que abrí el capítulo, he aquí la tragedia de los comunes del carbono. Los jugadores pueden ser ciudadanos individuales, cuya carga consiste en la inconveniencia de privarse de la carne, los viajes en avión o los vehículos utilitarios deportivos devoradores de gasolina. O pueden ser países enteros, en cuyo caso la carga son las repercusiones económicas de renunciar a la energía barata y portátil de los combustibles fósiles. Los números, como siempre, son arbitrarios, y la tragedia se plasma en su patrón: nos encaminamos a la celda inferior derecha.

		Todos los demás	
		Conservar	Emitir
Yo	Conservar	Carga −10 Carga −10	Cambio climático −100 Carga + Cambio climático −110
	Emitir	Carga −10 Beneficio +10	Cambio climático −100 Cambio climático −100

Al igual que un juramento vinculante puede evitar la traición mutua de los prisioneros en un dilema de dos personas, las leyes y los contratos vinculantes pueden castigar a las personas por el bien mutuo en un juego de bienes públicos. Un ejemplo puro es fácil de demostrar en un laboratorio. A un grupo de participantes se les entrega una suma de dinero y se les ofrece la posibilidad de contribuir a un fondo comunitario (el bien público), que luego el experimentador duplica y redistribuye. La mejor estrategia para todos es contribuir lo máximo, pero la mejor estrategia para cada individuo es guardarse su suma y dejar que contribuyan todos los demás. Los participantes captan la lúgubre lógica de la teoría de juegos y sus contribuciones se reducen a cero, a menos que se les dé, asimismo, la oportunidad de multar a los que van por libre, en cuyo caso las contribuciones se mantienen altas y todos ganan.

Fuera del laboratorio, un bien común en una comunidad en la que todos se conocen entre sí puede ser protegido mediante una versión multijugador del toma y daca: cualquier explotador de un recurso se convierte en blanco del chismorreo, la vergüenza, las amenazas veladas y el vandalismo discreto.[21] En comunidades más grandes y anónimas, las variaciones de los resultados han de hacerse mediante contratos y regulaciones vinculantes. De este modo pagamos impuestos para las carreteras, las escuelas y el sistema judicial, y los evasores van a la cárcel. Los ganaderos compran permisos de pastoreo y los pescadores respetan los

límites de sus capturas, siempre y cuando se les impongan también a los demás. Los jugadores de jóquey celebran las reglas de obligatoriedad del casco, que protege su cerebro sin ceder una ventaja de comodidad y visión a sus rivales. Y los economistas recomiendan un impuesto al carbono y la inversión en energías limpias, que reducen el beneficio privado de las emisiones y bajan el coste de la conservación, dirigiendo a todos hacia la recompensa común de la conservación mutua.

La lógica de los dilemas del prisionero y los bienes públicos socava el anarquismo y el libertarismo radical, pese al eterno encanto de la libertad sin restricciones. La lógica torna racional decir: «Debería existir una ley en contra de lo que estoy haciendo». Como decía Thomas Hobbes, el principio fundamental de la sociedad es «que un hombre esté dispuesto, cuando otros también lo estén [...], a renunciar a su derecho a todas las cosas, y se contente con tanta libertad contra otros hombres como consentiría a otros hombres contra él mismo».[22] Este contrato social no solo encarna la lógica moral de la imparcialidad. Nos libra asimismo de las tentaciones, los pagos del bobo y las tragedias de la traición mutua.

CAPÍTULO

9

Correlación y causalidad

Una de las primeras cosas que se enseñan en los manuales introductorios de Estadística es que la correlación no es lo mismo que la causalidad. Es también una de las primeras cosas que se olvidan.

THOMAS SOWELL[1]

La racionalidad abarca todas las esferas de la vida, incluida la personal, la política y la científica. No es de extrañar que los teóricos de la democracia estadounidense, inspirados por la Ilustración, fueran unos forofos de la ciencia, ni que los autócratas reales y potenciales se agarren a disparatadas teorías de causa y efecto.[2] Mao Zedong obligaba a los granjeros chinos a amontonar sus plantas de semillero para reforzar su solidaridad socialista, y un reciente líder estadounidense sugería que la COVID-19 podía tratarse con inyecciones de lejía.

Desde 1985 hasta 2006, Turkmenistán fue gobernado por el presidente de por vida Saparmurat Niyázov. Entre sus logros figuran convertir su autobiografía en lectura obligatoria para el examen de conducir de la nación y erigir una gigantesca estatua dorada de sí mismo que rotaba para mirar al sol. En 2004 publicó la siguiente notificación sanitaria a su devoto público: «Cuando era joven observaba a los perros jóvenes. Les daban huesos para roer. Aquellos de vosotros a los que se os han caído los dientes no roíais los huesos. Este es mi consejo».[3]

Dado que la mayoría de nosotros no corremos peligro alguno de ser encarcelados en Asjabad, podemos identificar el defecto del consejo de su excelencia. El presidente cometió uno de los más célebres errores en el razonamiento, confundiendo la correlación con la causalidad. Incluso si fuese cierto que los turkmenos desdentados no habían mordido huesos, el presidente no estaba autorizado a concluir que el hecho de roer huesos es lo que fortalece los dientes. Quizá solamente las personas con dientes fuertes puedan roer huesos, un caso de causalidad inversa. O tal vez algún tercer factor, como ser miembro del Partido Comunista, causara que los turkmenos royeran huesos (para demostrar lealtad a su líder) y tuvieran los dientes fuertes (si el cuidado dental fuese un requisito para ser miembro); un proceso confuso.

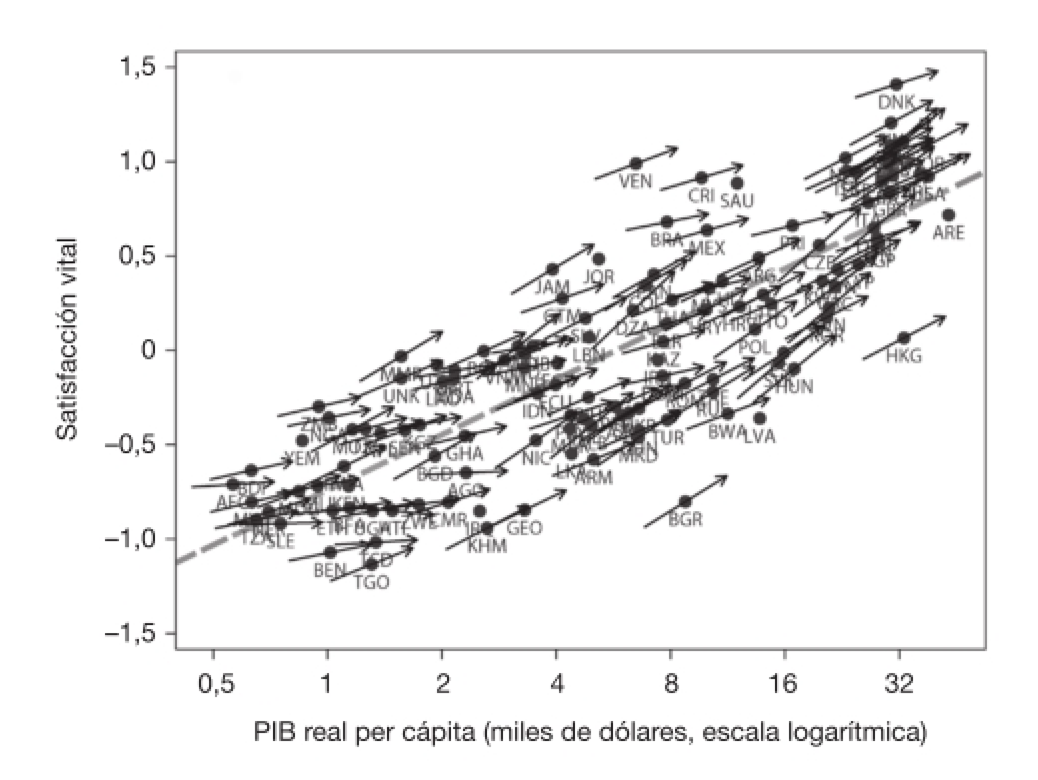
El concepto de *causalidad* y su contraste con la mera correlación es la savia de la ciencia. ¿Qué causa el cáncer? ¿O el cambio climático? ¿O la esquizofrenia? Está entretejido en nuestro lenguaje, razonamiento y humor ordinarios. El contraste semántico entre «el barco se hundió» y «el barco

fue hundido» estriba en si el hablante afirma que hubo un agente causal detrás del suceso en lugar de un acontecimiento espontáneo. Apelamos a la causalidad cada vez que ponderamos qué hacer con una gotera, una corriente o un dolor. Uno de los chistes favoritos de mi abuelo era sobre un hombre que se atiborró de *cholent* (el estofado de carne y judías cocido a fuego lento durante doce horas para almorzar en *sabbat*, cuando está prohibido cocinar) con un vaso de té, y luego yacía dolorido lamentándose de que el té le había sentado mal. Presumiblemente tendrías que haber nacido en Polonia en 1900 para que el chiste te pareciese tan desternillante como a él, pero si lo entiendes, puedes ver que la diferencia entre correlación y causalidad forma parte de nuestro sentido común.

No obstante, las confusiones niyázovianas son frecuentes en nuestro discurso público. Este capítulo investiga la naturaleza de la correlación, la naturaleza de la causalidad y las formas de distinguir ambas.

¿Qué es la correlación?

Una correlación es la dependencia del valor de una variable hacia el valor de otra: si conoces una, puedes predecir la otra, al menos aproximadamente (*predecir* significa aquí «suponer», no «pronosticar»; puedes predecir la estatura de los padres a partir de las estaturas de sus hijos o viceversa). Una correlación se representa con frecuencia en un gráfico denominado *diagrama de dispersión*. En el siguiente, cada punto representa un país, y los puntos están dispuestos de izquierda a derecha en función de sus ingresos medios, y arriba y abajo según su satisfacción vital autoevaluada (los ingresos se han comprimido en una escala logarítmica para compensar la utilidad marginal decreciente del dinero, por razones que vimos en el capítulo 6).[4]



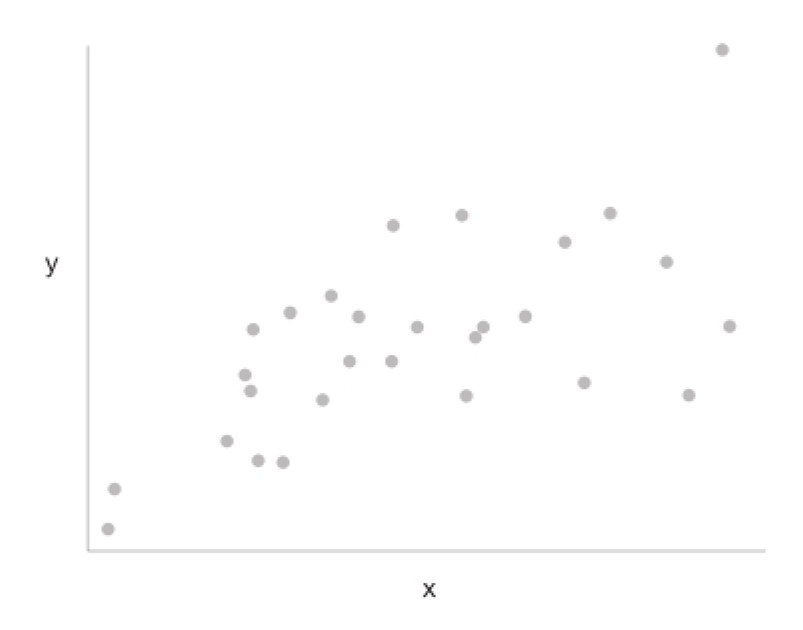
Adaptado con el permiso de Stevenson & Wolfers, 2008

Puedes captar inmediatamente la correlación: los puntos se extienden a lo largo de un eje diagonal, que se muestra como la línea gris discontinua que se oculta tras el enjambre. Cada punto es atravesado por una flecha que resume un minidiagrama de dispersión para la población *del* país. El macrodiagrama y los minidiagramas muestran que la felicidad está correlacionada con los ingresos, tanto entre los habitantes de un país (cada flecha) como en todos los países (los puntos). Y ya sé que estás resistiendo la tentación, al menos por ahora, de inferir que «la riqueza da la felicidad».

¿De dónde proceden la línea gris discontinua y las flechas que atraviesan cada punto? ¿Y cómo podríamos traducir nuestra impresión visual de que los puntos se extienden a lo largo de la diagonal a algo más objetivo, a fin de no engañarnos imaginando una raya con un montón de palitos chinos?

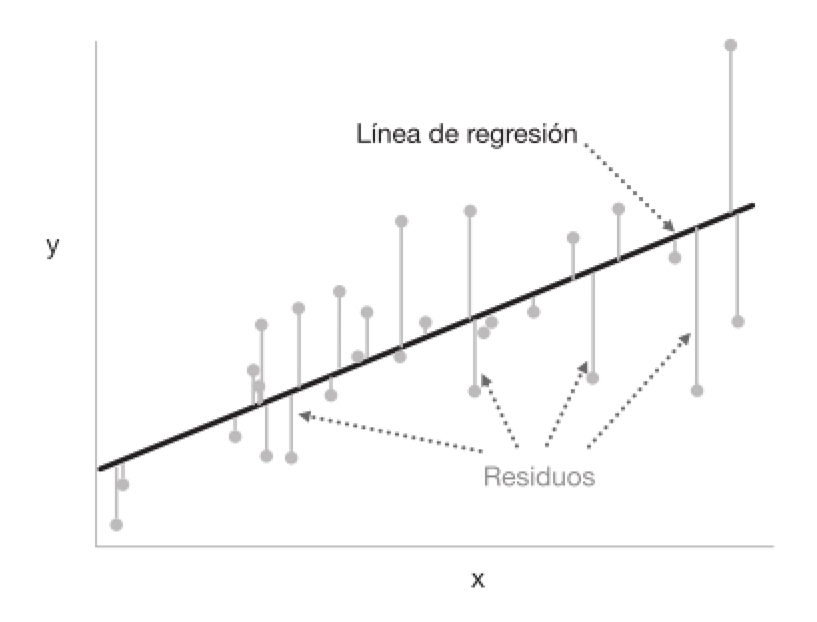
Esta es la técnica matemática llamada *regresión*, el caballo de batalla de la epidemiología y las ciencias sociales. Consideremos el diagrama de dispersión que hay más abajo. Imaginemos que cada punto de datos es una chincheta y que las conectamos a una varilla rígida con una goma elástica.

Imaginemos que las gomas solo pueden estirarse hacia arriba y hacia abajo, no en diagonal, y que cuanto más las estiramos, más resisten. Cuando todas las gomas están atadas, soltamos la varilla y dejamos que vuelvan a su lugar.

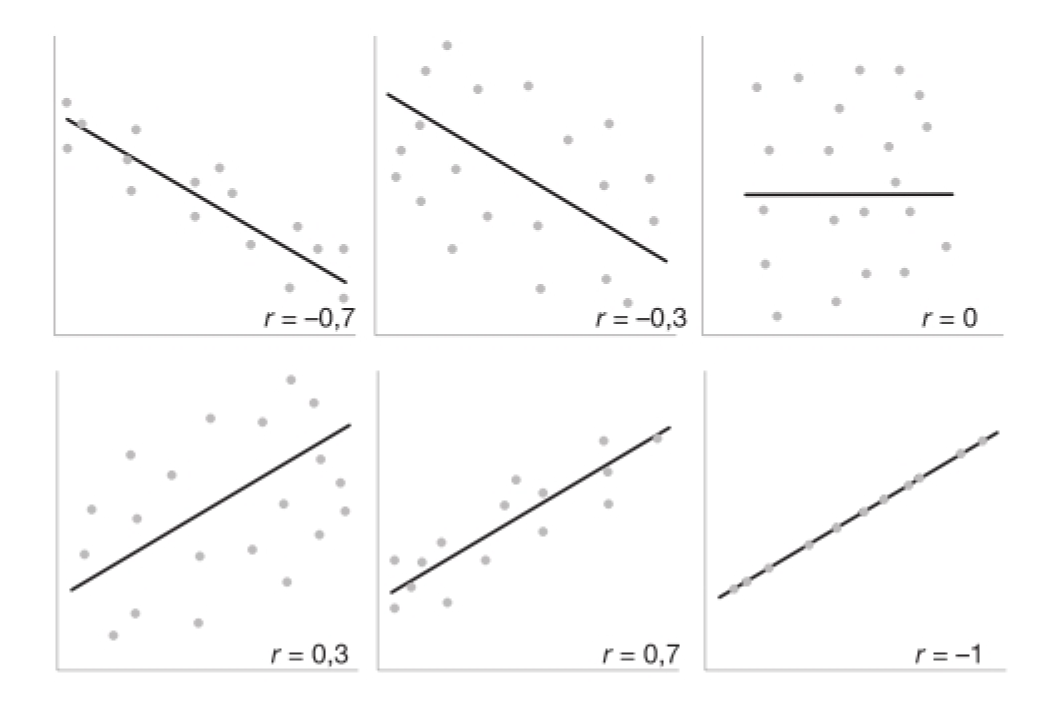


La varilla se instala en una posición y un ángulo que minimiza el cuadrado de la distancia entre cada chincheta y el sitio donde está atada. Así situada, la varilla se llama línea de regresión y captura la relación lineal entre las dos variables: *y*, que corresponde al eje vertical, y *x*, que corresponde al horizontal. La longitud de la goma elástica que conecta cada chincheta con la línea se denomina *residuo*, y capta la porción idiosincrásica del valor *y* de esa unidad que se resiste a ser predicha por su *x*. Regresemos al gráfico de la felicidad y los ingresos. Si la renta predijera perfectamente la felicidad, todos los puntos caerían exactamente a lo largo de la línea de regresión gris, pero, con los datos reales, eso jamás sucede. Algunos de los puntos flotan por encima de la línea (tienen grandes residuos positivos), como Jamaica, Venezuela, Costa Rica y Dinamarca. Dejando a un lado los errores de medición y otras fuentes de ruido, las discrepancias muestran que en 2006 (cuando se recogieron los datos), los habitantes de esos países eran más felices de lo que uno esperaría basándose en sus ingresos, quizá debido a otras características del país, como su clima

o su cultura. Otros puntos cuelgan por debajo de la línea, como Togo, Bulgaria y Hong Kong, lo cual sugiere que algo está haciendo que la gente de esos países esté un poco más triste de lo que le corresponde por su nivel de ingresos.



Los residuos nos permiten asimismo cuantificar el grado de correlación de las dos variables: cuanto más cortas las gomas, como proporción de cuánto se extiende el grupo entero de izquierda a derecha y de arriba abajo, más cerca de la línea están los puntos y más alta es la correlación. Con un poco de álgebra, esto puede convertirse en un número, *r*, el coeficiente de correlación, que oscila desde –1 (no mostrado), en cuya gráfica los puntos caen en fila a lo largo de una diagonal desde el noroeste hasta el sureste; pasando por un rango de valores negativos, donde salpican diagonalmente a lo largo de ese eje; pasando por 0, donde los puntos forman un enjambre no correlacionado de mosquitos; pasando por los valores positivos, donde salpican de suroeste a nordeste; hasta 1, donde los puntos se disponen perfectamente a lo largo de la diagonal.



Aunque las acusaciones de cometer errores al confundir la correlación con la causalidad suelen dirigirse contra aquellos que saltan de la primera a la segunda, a menudo el problema es más básico: no se ha establecido de entrada correlación alguna. Puede que los turkmenos que roan más huesos ni siquiera *tengan* los dientes más fuertes ($r = 0$). Los presidentes de las antiguas repúblicas soviéticas no son los únicos que no aciertan a mostrar la correlación, y menos aún la causalidad. En 2020 Jeff Bezos fanfarroneaba: «Todas mis mejores decisiones en los negocios y en la vida han sido tomadas con el corazón, la intuición, el instinto [...], no el análisis», dando a entender que el corazón y el instinto conducen a mejores decisiones que el análisis.[5] Pero no nos contó si todas sus *peores* decisiones en los negocios y en la vida *también* las tomaba con el corazón, la intuición y el instinto, ni si las buenas decisiones viscerales y las malas decisiones analíticas superaban en número a las malas decisiones viscerales y las buenas analíticas.

La *correlación ilusoria*, como se conoce esta falacia, se mostró por vez primera en un célebre conjunto de experimentos realizados por los psicólogos Loren y Jean Chapman, que se preguntaban por qué tantos psicoterapeutas continuaban usando el test de las manchas de tinta de Rorschach y el test del dibujo de una persona, aun cuando los estudios que habían tratado de validarlos no mostraban ninguna correlación entre las

respuestas a los test y los síntomas psicológicos. Los experimentadores emparejaron maliciosamente las descripciones de los pacientes psiquiátricos con respuestas en el test del dibujo, pero de hecho las descripciones eran falsas y los emparejamientos eran aleatorios. Luego pidieron a una muestra de estudiantes que refirieran cualquier patrón que detectaran en los pares.[6] Los estudiantes, guiados por sus estereotipos, estimaban incorrectamente que los hombres más anchos de hombros eran dibujados por pacientes hipermasculinos, los de ojos más abiertos eran obra de los paranoicos y así sucesivamente: exactamente los vínculos que los diagnosticadores profesionales aseguran detectar en sus pacientes, con el mismo poco fundamento en realidad.

Muchas correlaciones que han llegado a formar parte de nuestra sabiduría convencional, como que la gente acude en tropel a las urgencias de los hospitales durante una luna llena, son igualmente ilusorias.[7] El peligro es particularmente grave en las correlaciones que utilizan meses o años como unidades de análisis (los puntos del diagrama de dispersión), ya que muchas variables suben y bajan conjuntamente con los tiempos cambiantes. Un estudiante de Derecho aburrido, Tyler Vigen, diseñó un programa que hurga en la web en busca de conjuntos de datos con correlaciones carentes de sentido precisamente para mostrar cuán prevalentes son. El número de asesinatos por vapor u objetos calientes, por ejemplo, tiene una alta correlación con la edad de la actual Miss América. Y la tasa de divorcios en Maine sigue de cerca el consumo nacional de margarina.[8]

REGRESIÓN A LA MEDIA

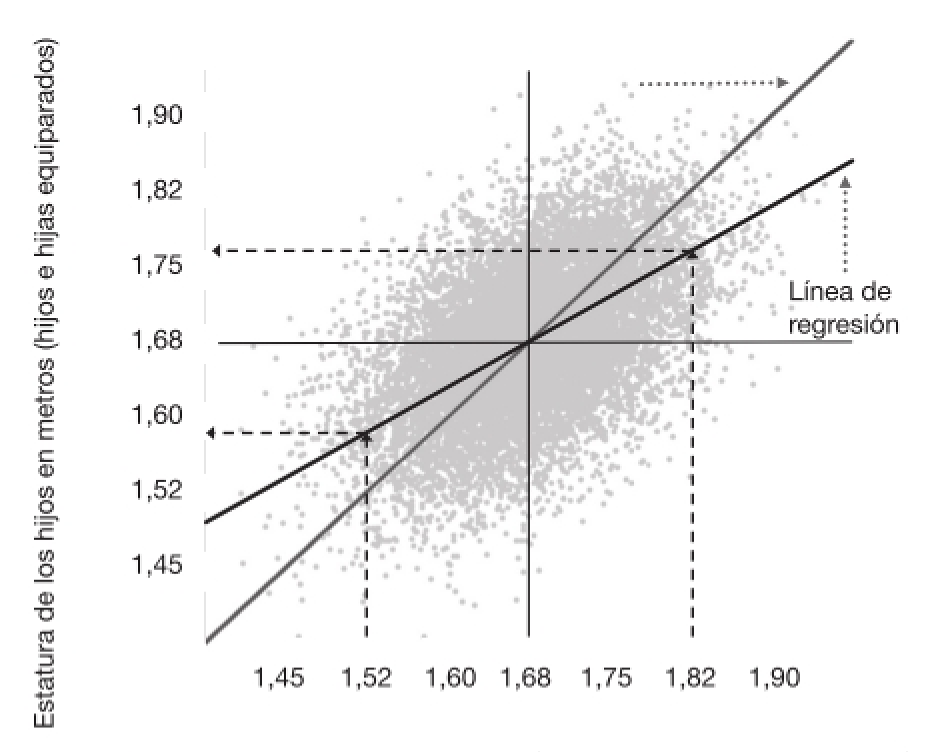
Regresión se ha convertido en el término estándar para los análisis correlacionales, pero la conexión es indirecta. El término hacía referencia originalmente a un fenómeno específico que acompaña a la correlación: la regresión a la media. Este omnipresente pero contraintuitivo fenómeno fue descubierto por el polímata victoriano Francis Galton (1822-1911), que representaba las estaturas de los hijos comparadas con la estatura media de sus dos padres (a mitad de camino entre la madre y el padre), ajustándola en

ambos casos para la diferencia media entre varones y mujeres. Descubrió que «cuando la estatura de los padres es superior a la media, sus hijos tienden a ser más bajos que ellos. Cuando la estatura de los padres es inferior a la media, sus hijos tienden a ser más altos que ellos».[9] Esto sigue siendo cierto, no solo para las estaturas de padres e hijos, sino también para el cociente intelectual (CI) de los padres y sus hijos, y, para el caso, para dos variables cualesquiera que no estén perfectamente correlacionadas. Un valor extremo de una tenderá a estar emparejado con un valor no tan extremo de la otra.

Esto no significa que las familias altas engendren hijos cada vez más bajos y viceversa, de suerte que algún día todos los niños medirán lo mismo y no habrá en el mundo yóqueis ni pívots. Tampoco significa que la población esté convergiendo en un CI medio de cien ni que se estén extinguiendo los genios y los zoquetes. La razón por la que las poblaciones no colapsan en la mediocridad uniforme, a pesar de la regresión a la media, es que las colas de la distribución se reabastecen constantemente con el ocasional hijo muy alto de padres más altos que la media y el hijo muy bajo de padres más bajos que la media.

La regresión a la media es un fenómeno puramente *estadístico,* una consecuencia del hecho de que, en las distribuciones en forma de campana, cuanto más extremo es un valor, menos probable es que aparezca. Eso implica que, cuando un valor es realmente extremo, cualquier otra variable que esté emparejada con él (como el hijo de una pareja muy alta) es improbable que esté a la altura de su rareza, o que repita su racha ganadora, o que reciba la misma mano afortunada, o que sufra la misma racha de mala suerte, o que capee la misma tormenta perfecta, una vez más, y recaerá hacia lo ordinario. En el caso de la estatura o del CI, la caprichosa conspiración sería cualquier combinación inusual de genes, experiencias y accidentes de la biología que confluyó en los padres. Muchos de los componentes de esa combinación serán favorecidos en sus hijos, pero la misma combinación no se reproducirá a la perfección (y viceversa: habida cuenta de que la regresión es un fenómeno estadístico, no causal, los padres también regresan a la media de sus hijos).

En una gráfica, cuando los valores correlacionados de dos curvas de campana se representen unos contra otros, el diagrama de dispersión se asemejará habitualmente a un balón de rugby inclinado. En la página siguiente tenemos un hipotético conjunto de datos similar al de Galton, que muestra las estaturas de los padres (la media de cada pareja) y las estaturas de sus hijos adultos (ajustadas de forma que los hijos y las hijas puedan representarse en la misma escala).



La diagonal gris de 45 grados muestra lo que esperaríamos como promedio si los hijos fueran exactamente tan excepcionales como sus padres. La línea de regresión negra es lo que encontramos en realidad. Si nos centramos en un valor extremo, por ejemplo, en los padres con una media de estatura entre ellos de 1,90 metros, descubriremos que la agrupación de puntos que representa a sus hijos cuelga básicamente por debajo de la diagonal de 45 grados, como podemos confirmar recorriendo la flecha punteada derecha hasta la línea de regresión, girando hacia la izquierda y siguiendo la flecha punteada horizontal hasta el eje vertical, donde señala un poco por encima de 1,75 metros, siendo más bajos que sus

padres. Si nos fijamos en los padres con una estatura media de 1,52 metros (flecha punteada izquierda), veremos que sus hijos flotan básicamente por encima de la diagonal gris, y el giro a la izquierda en la línea de regresión nos conduce a un valor de casi 1,60 metros, siendo más altos que sus padres.

La regresión a la media sucede cada vez que dos variables se correlacionan imperfectamente, lo cual significa que tenemos una larga experiencia acumulada en ella. No obstante, Tversky y Kahneman han demostrado que la mayoría de las personas no son conscientes del fenómeno (a pesar del chiste malo de la viñeta de *Frank y Ernest*, que juega con el doble sentido del término inglés *mean*, que significa tanto «media» como «malo»).[10]



Frank and Ernest. Utilizado con el permiso de Thaves and the Cartoonist Group.

Los individuos dirigen su atención hacia un acontecimiento porque este es inusual, y no aciertan a prever que cualquier cosa asociada con ese suceso probablemente no será tan inusual como este. Antes bien, proponen explicaciones causales falaces para lo que de hecho es una inevitabilidad estadística.

Un ejemplo trágico es la ilusión de que la crítica funciona mejor que el elogio y el castigo mejor que la recompensa.[11] Criticamos a los estudiantes cuando obtienen malos resultados. Pero es improbable que la mala suerte que maldijo esos resultados se repita en el próximo intento, por lo que seguro que mejorarán, con lo que nos engañamos pensando que el castigo funciona. Los elogiamos cuando hacen bien las cosas, pero el rayo nunca

cae dos veces en el mismo lugar, así que resulta improbable que repitan esa hazaña la próxima vez, por lo que nos engañamos pensando que el elogio es contraproducente.

El desconocimiento de la regresión a la media abona el terreno para muchas otras ilusiones. Los aficionados a los deportes teorizan sobre por qué el Rookie del Año* está condenado a sufrir una depresión de segundo año, y por qué el tema de portada de una famosa revista tendrá que sufrir la maldición de *Sports Illustrated* (¿exceso de confianza?, ¿expectativas imposibles?, ¿las distracciones de la fama?). Pero si un atleta destaca por una semana o un año extraordinario, resulta improbable que las estrellas se alineen de esa forma dos veces consecutivas, y no puede sino regresar hacia la media (con la misma falta de sentido, un equipo en caída libre mejorará cuando despidan al entrenador). Después de que una ola de crímenes horrendos inunde todos los periódicos, los políticos intervienen con equipos especiales de intervención policial, equipamiento militar, carteles de vigilancia vecinal y otros trucos efectistas y, por descontado, al mes siguiente se felicitan a sí mismos porque el índice de criminalidad ya no es tan elevado. También los psicoterapeutas, con independencia de su modalidad de «cura por la palabra», pueden declarar una victoria no ganada después de tratar a un paciente que acude con un ataque de ansiedad o depresión severa.

Una vez más, los científicos no son inmunes. Otra de las causas de los fallos de replicación es que los experimentadores no aprecian una versión de la regresión a la media denominada *maldición del ganador*. Si los resultados de un experimento parecen demostrar un efecto interesante, muchas cosas deben de haber salido bien, tanto si el efecto es real como si no lo es. Los dioses del azar deben de haber sonreído a los experimentadores, que no deberían contar con ello una segunda vez, por lo que, cuando traten de replicar el efecto, deberían reclutar *más* participantes. Pero la mayoría de los experimentadores creen haber acumulado ya ciertas evidencias del efecto, por lo que pueden permitirse *menos* participantes, sin percatarse de que esta estrategia es un camino de sentido único hacia el *Journal of Irreproducible Results [Revista de Resultados Irreproducibles]*.[12] La incapacidad de apreciar que la regresión a la media

es aplicable a los descubrimientos extraordinarios condujo a un confuso artículo publicado en 2010 en *The New Yorker* con el título «The Truth Wears Off» [«La verdad desaparece»], que postulaba un *efecto de declive*, que supuestamente arrojaba dudas sobre el método científico.[13]

La maldición del ganador es aplicable a cualquier empresa humana inusualmente exitosa, y nuestra incapacidad de compensar los momentos singulares de buena fortuna puede ser una de las razones por las que la vida trae consigo a menudo la decepción.

¿QUÉ ES LA CAUSALIDAD?

Antes de tender el puente desde la correlación hasta la causalidad, examinemos la otra orilla, la *causalidad* misma. Este resulta un concepto sorprendentemente escurridizo.[14] Una vez más, Hume estableció el marco para siglos de análisis al aventurar que la causalidad es meramente una expectativa de que una correlación que hemos experimentado en el pasado se mantendrá en el futuro.[15] Una vez que hemos observado suficientes billares, cada vez que vemos una bola acercarse a una segunda, prevemos que la segunda saldrá despedida, como sucedió en todas las ocasiones anteriores, apoyándonos en nuestra asunción tácita, pero indemostrable, de que las leyes de la naturaleza persisten a lo largo del tiempo.

No se tarda mucho en detectar la deficiencia de la *conjunción constante* como una teoría de la causalidad. El gallo canta siempre justo antes del alba, pero no le atribuimos el mérito de causar la salida del sol. Análogamente, los truenos preceden con frecuencia los incendios forestales, pero no decimos que los truenos causen los incendios. Estos son *epifenómenos*, conocidos también como *factores de confusión* o *variables estorbo*: acompañan al suceso, pero no lo causan. Los epifenómenos son la pesadilla de la epidemiología. Durante muchos años se culpó al café de las enfermedades cardiacas porque los bebedores de café sufrían más ataques al corazón. Resultó que los bebedores de café tienden asimismo a fumar y a evitar el ejercicio; el café era un epifenómeno.

Hume anticipó el problema y elaboró su teoría: no solo la causa tiene que preceder habitualmente a su efecto, sino que, además, «si el primer objeto no hubiera estado presente, el segundo nunca habría existido». La cláusula crucial «si no hubiera estado presente» es un *contrafáctico*, un *y si...* Se refiere a lo que ocurriría en un mundo posible, un universo alternativo, un experimento hipotético. En un universo paralelo en el que la causa no sucediera, tampoco existiría el efecto. La definición contrafáctica de la causalidad resuelve el problema de los epifenómenos. La razón por la que decimos que el gallo no causa el amanecer es que *si* el gallo se hubiera convertido en el ingrediente principal del *coq au vin* la noche anterior, el sol habría salido igualmente. Decimos que los rayos causan los incendios forestales y los truenos no porque, si hubiera rayos sin truenos, un bosque podría incendiarse, pero no viceversa.

Por tanto, la causalidad puede concebirse como la diferencia entre los resultados cuando tiene lugar un suceso (la causa) y cuando este no se produce.[16] El *problema fundamental de la inferencia causal*, como lo denominan los estadísticos, estriba en que estamos atrapados en este universo, donde o bien un supuesto suceso causal tuvo lugar o bien no se produjo. No podemos asomarnos a ese otro universo para ver cuál es allí el resultado. Sin duda, podemos comparar los resultados en este universo en las diversas ocasiones en las que esa clase de suceso tiene o no lugar. Pero eso choca de lleno contra un problema señalado por Heráclito en el siglo VI a. C. No puedes meterte dos veces en el mismo río. Entre esas dos ocasiones, el mundo puede haber cambiado en otros sentidos, y no puedes estar seguro de si uno de esos otros cambios era la causa. Podemos comparar asimismo las cosas concretas que estuvieron sometidas a esa clase de evento con cosas similares que no lo estuvieron. Pero esto también se topa con un problema, señalado por Dr. Seuss: «Hoy tú eres tú, eso es más cierto que lo cierto. No hay nadie vivo que sea más tú que tú». Cada individuo es único, por lo que no podemos saber si un resultado experimentado por un individuo dependía de la supuesta causa o bien de la miríada de particularidades de esa persona. Para inferir la causalidad a partir de estas comparaciones, hemos de suponer, como se dice menos

poéticamente, la *estabilidad temporal* y la *homogeneidad de las unidades*. Los métodos examinados en las dos secciones siguientes tratan de tornar razonables esos supuestos.

Incluso una vez que hemos establecido que una causa marca una diferencia en un resultado, ni científicos ni legos se conforman con dejar así las cosas. Conectamos la causa con su efecto con un *mecanismo*: el mecanismo de relojería entre bastidores que empuja las cosas. Tenemos la intuición de que el mundo no es un videojuego con patrones de píxeles que dan paso a nuevos patrones. Por debajo de cada acontecimiento hay una fuerza, un poder o un brío oculto. A la luz de la ciencia, muchas de nuestras intuiciones primitivas acerca de los poderes causales resultan ser erróneas, como el «ímpetu» que los medievales creían que estaba impreso en los objetos móviles, y el *psi*, el *qi*, los engramas, los campos de energía, los miasmas homeopáticos, el poder de los cristales y demás bobadas de la medicina alternativa. Pero ciertos mecanismos intuitivos, como la gravedad, sobreviven en formas científicamente respetables. Y se han postulado muchos nuevos mecanismos ocultos para explicar correlaciones en el mundo, entre los que figuran los genes, los patógenos, las placas tectónicas y las partículas elementales. Estos mecanismos causales son los que nos permiten predecir lo que sucedería en los escenarios contrafactuales, levantándolos del reino de la fantasía: construimos el mundo imaginario y luego simulamos los mecanismos que operan en él.

INCLUSO SI CONCEBIMOS la causación en términos de los resultados alternativos y los mecanismos que los producen, cualquier esfuerzo para identificar *la* causa de un efecto plantea una maraña de enigmas. Uno de ellos es la elusiva diferencia entre *causa* y *condición*. Decimos que encender una cerilla causa un fuego, porque sin prenderla no habría fuego. Pero sin oxígeno, sin la sequedad del papel, sin la quietud de la habitación, tampoco habría ningún fuego. ¿Por qué no decimos entonces que «el oxígeno causó el fuego»?

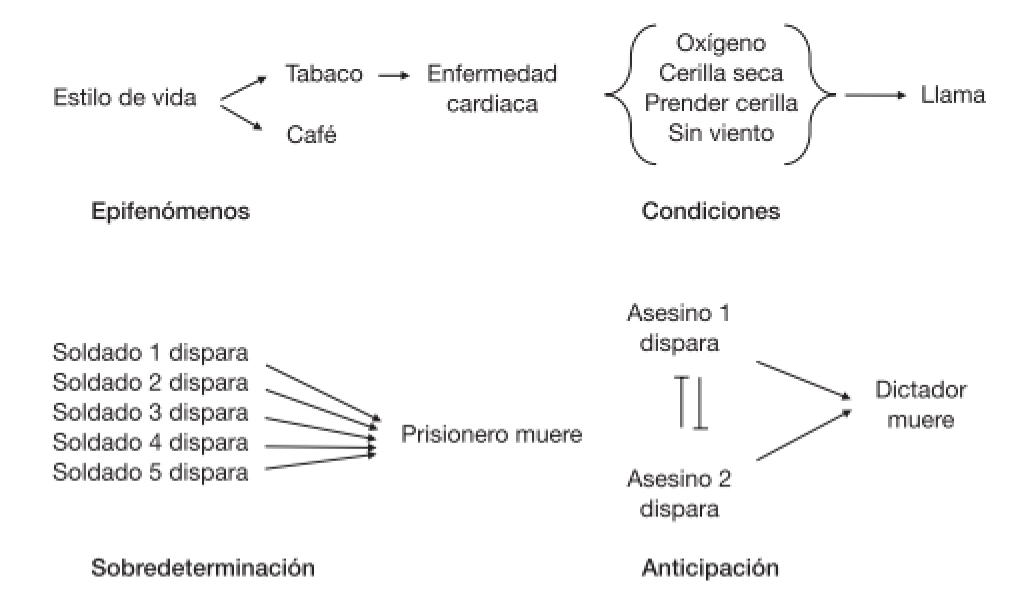
Un segundo enigma es la *anticipación*. Supongamos, a título de ejemplo, que Lee Harvey Oswald hubiera tenido un compañero de conspiración en la verde colina de Dallas en 1963, y ambos hubieran

planeado que el primero que tuviese a tiro a JFK le dispararía, mientras el otro se mezclaba con la multitud. En el mundo contrafáctico en el que Oswald no hubiese disparado, JFK habría muerto de todos modos, pero resultaría descabellado negar que, en el mundo en el que disparó antes que su cómplice, él causó la muerte de Kennedy.

El tercero es la *sobredeterminación*. Un prisionero condenado a muerte es ejecutado por un pelotón de fusilamiento en lugar de por un solo verdugo, a fin de que ningún tirador tenga que vivir con la terrible carga de ser quien causó la muerte: si él no hubiera disparado, el prisionero habría muerto de todas formas. Pero entonces, en virtud de la lógica de los contrafácticos, *nadie* causó su muerte.

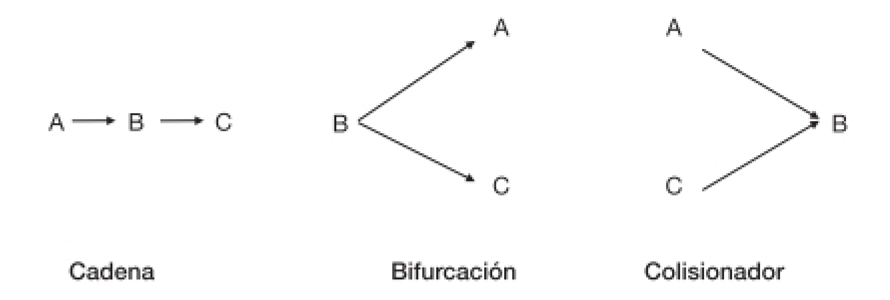
Y luego está la *causalidad probabilística*. Muchos de nosotros conocemos a algún nonagenario que fumó un paquete diario toda su vida. Ahora bien, hoy en día pocos dirían que su avanzada edad demuestra que fumar no causa cáncer, aunque esa era una «refutación» habitual en los tiempos anteriores a que se volviese innegable el vínculo entre el tabaco y el cáncer. Incluso en la actualidad es muy habitual la confusión entre causalidad imperfecta y no causalidad. Un editorial publicado en 2020 en *The New York Times* abogaba en favor de la abolición de la policía porque «la estrategia actual no ha terminado con [la violación]. La mayoría de los violadores nunca ven el interior de una sala de tribunal».[17] El editorialista no consideraba si, en caso de que no hubiera policía, aún menos violadores, o ninguno en absoluto, verían el interior de una sala de tribunal.

Podemos comprender estas paradojas de la causalidad olvidando simplemente las bolas de billar y reconociendo que ningún suceso tiene una única causa. Los acontecimientos están insertos en una *red* de causas que se desencadenan, se posibilitan, se inhiben, se impiden y se sobrealimentan unas a otras por sendas conectadas y ramificadas. Los cuatro enigmas causales devienen menos enigmáticos cuando extendemos en cada caso el mapa de carreteras de la causalidad mostrado a continuación.



Si interpretamos las flechas no como implicaciones lógicas («Si *x* fuma, entonces *x* padece una enfermedad cardiaca», sino como probabilidades condicionales («la verosimilitud de que *x* padezca una enfermedad cardiaca dado que *x* es un fumador es más alta que la verosimilitud de que *x* padezca una enfermedad cardiaca dado que no es un fumador»), y los nodos de eventos no como encendidos o apagados, sino como probabilidades que reflejan una tasa base o probabilidad *a priori*, entonces el diagrama se denomina *red bayesiana causal*.[18] Podemos calcular lo que se desarrolla a lo largo del tiempo aplicando (naturalmente) la regla de Bayes nodo a nodo a través de la red. Por muy enrevesada que sea la maraña de causas, condiciones y confusiones, podemos determinar entonces qué sucesos son causalmente dependientes o independientes entre sí.

El inventor de estas redes, el científico informático Judea Pearl, observa que estas se construyen a partir de tres sencillos patrones (la cadena, la bifurcación y el colisionador), cada uno de los cuales captura una característica fundamental (pero no intuitiva) de la causalidad con más de una causa.



Las conexiones reflejan las probabilidades condicionales. En cada caso, A y C no están conectadas directamente, lo cual significa que la probabilidad de A, dada B, puede especificarse independientemente de la probabilidad de C, dada B. Y en cada caso puede decirse algo distintivo de la relación entre ellas.

En una *cadena* causal, la primera causa, A, se «separa» del efecto final, C, su única influencia es a través de B. Por lo que atañe a C, A podría no existir. Consideremos la alarma contra incendios de un hotel, activada por la cadena «fuego → humo → alarma». En realidad no es una alarma contra el fuego, sino una alarma contra el humo; de hecho, es una alarma contra la neblina. Los huéspedes pueden ser despertados con la misma facilidad por alguien que esté pintando con aerosol una estantería cerca de una entrada de ventilación como por alguien que esté flameando una *crème brûlée*.

Una *bifurcación causal* es algo familiar: representa una confusión o epifenómeno, con el peligro concomitante de identificar erróneamente la causa real. La edad (B) influye en el vocabulario (A) y en el número del calzado (C), ya que los niños mayores tienen los pies más grandes y conocen más palabras. Esto significa que el vocabulario está correlacionado con el número del calzado. Pero sería poco aconsejable que el programa de aprendizaje temprano Head Start preparase a los niños para la escuela calzándolos con unas deportivas más grandes.

Igual de peligroso resulta el *colisionador*, donde las causas no relacionadas convergen en un único efecto. En realidad, esto es más peligroso todavía, porque, aunque la mayoría de la gente detecta intuitivamente la falacia de una confusión (se partían de risa en el *shtetl* con

el chiste del *cholent* y el vaso de té), el *sesgo de selección de estratificación del colisionador* es prácticamente desconocido. La trampa en un colisionador causal radica en que, cuando nos centramos en un rango restringido de efectos, introducimos una correlación negativa artificial entre las causas, toda vez que una causa compensará la otra. Muchos veteranos del mundo de las citas se preguntan por qué los hombres atractivos son idiotas. Pero puede tratarse de una calumnia contra los guapos, y es una pérdida de tiempo inventar teorías para explicarla, como que los hombres atractivos se han echado a perder porque han sido adulados durante toda su vida. Muchas mujeres saldrán con un hombre (B) solo si este es atractivo (A) o simpático (C). Incluso si la simpatía y el aspecto no estuvieran correlacionados en el mundo de las citas, los hombres menos agraciados *tendrían* que ser simpáticos o la mujer nunca habría salido con ellos, mientras que los cachas no eran clasificados mediante ningún filtro semejante. Una falsa correlación negativa fue introducida por su quisquillosidad disyuntiva.

La falacia del colisionador engaña a los críticos de las pruebas estandarizadas haciéndoles creer que los resultados de las pruebas no importan, basándose en la observación de que los estudiantes de posgrado que fueron admitidos con puntuaciones más altas no tienen más probabilidades de completar el programa. El problema estriba en que los estudiantes que fueron aceptados *a pesar* de sus bajas puntuaciones debían de poseer *otros* valores.[19] Si uno no es consciente del sesgo, podría concluir incluso que el tabaquismo materno es bueno para los bebés, ya que entre los bebés con poco peso al nacer, los hijos de madres fumadoras están más sanos. Ello se debe a que el poco peso al nacer ha de ser causado por *algo*, y las otras causas posibles, como el alcoholismo o la drogadicción, pueden ser más nocivas aún para el hijo.[20] La falacia del colisionador explica asimismo por qué Jenny Cavilleri[*] mantenía injustamente que los muchachos ricos son estúpidos: para entrar en Harvard (B) puedes ser o bien rico (A) o bien inteligente (C).

De la correlación a la causalidad: experimentos reales y naturales

Ahora que hemos investigado la naturaleza de la correlación y la naturaleza de la causalidad, ha llegado el momento de ver cómo llegar de la una a la otra. El problema no es que «correlación no implica causalidad». Normalmente lo hace porque, a menos que la correlación sea ilusoria o una mera coincidencia, *algo* debe haber causado que una variable se alinee con la otra. El problema reside en que, cuando una cosa está correlacionada con otra, eso no significa necesariamente que la primera causara la segunda. Como dice el mantra: cuando A está correlacionado con B, eso podría significar que A causa B, que B causa A o que algún tercer factor, C, causa tanto A como B.

La causación inversa y la confusión, versos segundo y tercero del mantra, son omnipresentes. El mundo es una inmensa red bayesiana causal, con flechas que apuntan en todas direcciones, entrelazando los sucesos en nudos en los que todo está correlacionado con todo lo demás. Estos nudos (llamados *multicolinealidad* y *endogeneidad*) pueden surgir en virtud del efecto Mateo, sucintamente explicado por Billie Holiday: «Los que tienen recibirán, los que no tienen perderán. Eso decía la Biblia y aún es novedad».[21*] Los países que son más ricos tienden a ser también más saludables, más felices, más seguros, más educados, menos contaminados, más pacíficos, más democráticos, más liberales, más laicos y más igualitarios en cuestión de género.[22] Las personas más ricas tienden a ser, asimismo, más saludables, más educadas y a estar mejor relacionadas, y tienen más probabilidades de hacer ejercicio, comer bien y pertenecer a grupos privilegiados.[23]

Estas marañas significan que prácticamente cualquier conclusión causal que saquemos de las correlaciones entre países o entre personas es probable que sea errónea o, en el mejor de los casos, no demostrada. ¿Hace la democracia más pacífico un país porque su líder no puede convertir fácilmente a los ciudadanos en carne de cañón? ¿O los países que no se enfrentan a amenazas de sus vecinos pueden permitirse el lujo de la democracia? ¿El hecho de ir a la universidad te equipa con destrezas que te permiten ganarte bien la vida? ¿O solamente las personas inteligentes, disciplinadas o privilegiadas, que son capaces de traducir sus valores naturales en activos financieros, consiguen ir a la universidad?

Existe una manera impecable de deshacer estos nudos: el experimento aleatorio, denominado con frecuencia *ensayo controlado aleatorio* o ECA. Tomemos una amplia muestra de la población de interés, aleatoriamente dividida en dos grupos, apliquemos la supuesta causa a un grupo y al otro no, y veamos si el primer grupo cambia y el segundo no. El experimento aleatorio es lo más que podemos acercarnos a la *creación* del mundo contrafáctico que es la prueba de fuego de la causalidad. En una red causal, este consiste en separar quirúrgicamente la supuesta causa de todas las influencias recibidas, estableciendo diferentes valores, y ver si difieren las probabilidades de los supuestos efectos.[24]

La aleatoriedad es la clave: si los pacientes a quienes se suministró el medicamento se inscribieron antes, o vivían más cerca del hospital, o tenían más síntomas interesantes, que los pacientes que ingirieron el placebo, nunca sabremos si el fármaco funcionó. Como decía uno de mis profesores de posgrado (aludiendo a una frase de la obra teatral de J. M. Barrie *What Every Woman Knows [Lo que todas las mujeres saben]*): «La asignación aleatoria es como el encanto. Si lo tienes, no necesitas nada más; si no lo tienes, no importa qué otras cosas tengas».[25] Esto no es del todo cierto en lo que concierne al encanto, ni es del todo cierto tampoco con respecto a la asignación aleatoria, pero todavía me acompaña décadas después, y lo prefiero al tópico de que los ensayos aleatorios son el «patrón oro» para demostrar la causalidad.

La sabiduría de los ensayos controlados aleatorios está penetrando en la política, la economía y la educación. Cada vez más, sus paladines están instando a los responsables políticos a poner a prueba sus panaceas en un conjunto de pueblos, clases o barrios aleatoriamente seleccionados, y comparar los resultados con un grupo de control que se pone en una lista de espera o al que se le asigna un programa de poca monta para mantenerlo ocupado.[26] Es probable que el conocimiento obtenido logre mejores resultados que las formas tradicionales de evaluar las políticas, como el dogma, el folclore, el carisma, la sabiduría convencional y la opinión de las personas mejor pagadas.

Los experimentos aleatorios no son ninguna panacea (pues nada es una panacea, lo cual es una buena razón para retirar ese cliché). Los científicos de laboratorio se critican unos a otros tanto como los científicos de datos correlacionales, porque incluso en un experimento no puedes hacer una sola cosa. Los experimentadores pueden pensar que han administrado un tratamiento y solo ese tratamiento al grupo experimental, pero otras variables pueden confundirse con este, un problema denominado *excluibilidad*. Cuenta un chiste que una pareja sexualmente insatisfecha consulta con un rabino su problema, pues está escrito en el Talmud que el marido es responsable del placer sexual de su esposa. El rabino se acaricia la barba y propone una solución: deberían contratar a un joven fornido y atractivo para que agite una toalla sobre ellos la próxima vez que hagan el amor, y las fantasías ayudarán a la mujer a alcanzar el clímax. Siguen el consejo del gran sabio, pero no consiguen el efecto deseado, y le suplican orientación una vez más. Él se acaricia la barba y se le ocurre una variación. En esta ocasión, el joven hará el amor a la mujer y el marido agitará la toalla. Siguen su consejo y, por supuesto, la mujer disfruta de un orgasmo extático y sísmico. El marido le dice al hombre: «¡Imbécil! *Así* es como se agita una toalla».

Por supuesto, el otro problema de las manipulaciones experimentales es que el mundo no es un laboratorio. Los politólogos no pueden lanzar una moneda al aire, imponer la democracia en unos países y la autocracia en otros, y esperar cinco años para ver cuáles van a la guerra. Los mismos problemas prácticos y éticos afectan a los estudios de los individuos, como se muestra en esta viñeta.



«El título de mi proyecto de ciencias es *Mi hermano pequeño: naturaleza o crianza*.»
Michael Shaw/The New Yorker Collection/The Cartoon Bank

Aunque no todo puede ser estudiado en un ensayo experimental, los científicos sociales han hecho acopio de ingenio para hallar casos en los que el mundo haga la aleatorización por ellos. Estos experimentos de la naturaleza pueden permitirnos a veces extraer conclusiones causales de un universo correlacional. Estos son un elemento recurrente en *Freakonomics*, la serie de libros y otros medios del economista Steven Levitt y el periodista Stephen Dubner.[27]

Un ejemplo es la *discontinuidad de la regresión*. Supongamos que quieres decidir si ir a la universidad hace más ricas a las personas o si los adolescentes destinados a la riqueza tienen más probabilidades de entrar en la universidad. Aunque no puedes aleatorizar literalmente una muestra de adolescentes y obligar a una universidad a admitir a un grupo y rechazar a otro, las universidades selectivas lo hacen eficazmente con los estudiantes que están cerca de su nota de corte. Nadie cree realmente que el estudiante que aprobó por los pelos con una puntuación en la prueba de 1720 es más inteligente que el que se quedó a las puertas con 1710. La diferencia está en el ruido, y podría haber sido igualmente aleatoria (lo mismo sucede con otras cualificaciones, como las notas y las cartas de recomendación).

Supongamos que seguimos a ambos grupos durante una década y representamos sus ingresos comparándolos con las respectivas puntuaciones de sus pruebas. Si vemos un escalón o un codo en la nota de corte, con un salto mayor en el salario en el límite entre el rechazo y la admisión que para intervalos de tamaño similar a lo largo del resto de la escala, podemos concluir que la varita mágica de la admisión marcó una diferencia.

Otro regalo para los científicos sociales hambrientos de causalidad es la aleatorización fortuita. ¿El canal de noticias Fox News vuelve más conservadora a la gente o los conservadores gravitan hacia Fox News? Cuando Fox News debutó en 1996, diferentes compañías de cable lo incorporaron caprichosamente a su oferta durante los cinco años siguientes. Los economistas sacaron partido de la casualidad durante medio decenio y descubrieron que en las localidades con Fox News en su oferta de cable se votaba entre 0,4 y 0,7 puntos más a los republicanos que en las localidades que tenían que ver otros canales.[28] Esa diferencia es suficientemente grande para decidir unas elecciones reñidas, y el efecto podría haberse acumulado en las décadas subsiguientes, cuando la penetración universal de Fox News en los mercados televisivos volvió el efecto más difícil de demostrar, pero no menos potente.

Más difícil, pero no imposible. Otra genialidad responde al poco útil nombre de *regresión variable instrumental*. Supongamos que deseamos ver si A causa B y nos preocupan las molestias habituales de la causalidad inversa (B causa A) y la confusión (C causa A y B). Supongamos ahora que encontramos una cuarta variable, I (el «instrumento»), que está correlacionada con la supuesta causa, A, pero no podría ser causada por ella, por ejemplo, porque sucedió antes en el tiempo y el futuro es incapaz de influir en el pasado. Imaginemos asimismo que esta variable prístina tampoco está correlacionada con la confusión, C, y que no puede causar directamente B, sino solo a través de A. Aun cuando A no pueda asignarse aleatoriamente, tenemos la segunda mejor opción, I. Si I, la clara sustituta de A, resulta estar correlacionada con B, eso es una indicación de que A causa B.

¿Qué tiene que ver esto con Fox News? Otro regalo para los científicos sociales es la pereza estadounidense. Los estadounidenses odian bajarse del coche, añadir agua a una mezcla para sopa y hacer clic en el menú de los canales por cable de más de un dígito. Cuanto más bajo el número del canal, más gente lo ve. Pues bien, las diferentes compañías de televisión por cable asignan a Fox News distintos números de canales al azar (la numeración dependía únicamente de cuándo la cadena cerraba un trato con cada empresa de cable, y no guardaba relación con los datos demográficos de los espectadores). Mientras que un número de canal bajo (I) puede hacer que la gente vea Fox News (A), y ver Fox News puede o no hacerles votar a los republicanos (B), ni tener ideas conservadoras (C) ni votar a los republicanos puede causar que la cadena de televisión favorita de alguien baje puestos en los canales por cable. En efecto, en una comparación entre los mercados del cable, cuanto más bajo es el número de canal de Fox News respecto a otras cadenas de noticias, mayor es el voto al Partido Republicano.[29]

De la correlación a la causalidad sin experimentación

Cuando un científico de datos encuentra una discontinuidad de la regresión o una variable instrumental, es una estupenda noticia. Pero es más frecuente que tenga que sacar la causalidad que pueda de la habitual maraña correlacional. No todo está perdido, sin embargo, ya que existen paliativos para cada una de las dolencias que debilitan la inferencia causal. Estos no son tan buenos como el encanto de la asignación aleatoria, pero al menos son lo mejor que podemos hacer en un mundo que no fue creado en beneficio de los científicos.

La causalidad inversa es la más fácil de descartar de las dos, gracias a la ley de hierro que constriñe a los escritores de ciencia ficción y otras tramas de viajes en el tiempo como *Back to the Future (Regreso al futuro)*: el futuro no puede influir en el pasado. Supongamos que queremos poner a prueba la hipótesis de que la democracia causa la paz, no solo viceversa. Lo primero es evitar la falacia de la causalidad de todo o nada, e ir más allá de la afirmación común, pero falsa, de que «las democracias nunca luchan

entre sí» (hay infinidad de excepciones).[30] La hipótesis más realista es que los países que son *relativamente* más democráticos tienen *menos probabilidades* de caer en una guerra.[31] Varios organismos de investigación dan a los países puntuaciones desde –10 para una autocracia plena, como Corea del Norte, hasta +10 para una democracia plena, como Noruega. Con la paz resulta un poco más difícil, porque (por fortuna para la humanidad, pero por desgracia para los científicos sociales) las guerras abiertas son infrecuentes, por lo que la mayoría de las entradas de la tabla serían cero. En vez de ello, podemos calcular la predisposición a la guerra en función del número de «disputas militarizadas» en las que un país ha estado envuelto durante más de un año: alardes belicosos, estado de alerta de las fuerzas armadas, disparos de advertencia, envíos de aviones de combate, amenazas belicistas y escaramuzas fronterizas. Podemos convertir esto de una puntuación de la guerra en una puntuación de la paz (de modo de los países más pacíficos consigan números más altos), restando el cómputo a un número mayor, como el número máximo de disputas jamás registradas. Acto seguido podemos correlacionar la puntuación de la paz con la puntuación de la democracia. Por supuesto, por sí sola esa correlación no demuestra nada.

Ahora bien, supongamos que cada variable se registra *dos veces*, por ejemplo, con una década de distancia. Si la democracia causa la paz, entonces la puntuación de la democracia en el momento 1 debería estar correlacionada con la puntuación de la paz en el momento 2. Esto tampoco demuestra gran cosa, porque en una década el leopardo no puede cambiar sus manchas: entonces, una democracia pacífica solo puede ser una democracia pacífica ahora. Pero, a modo de control, podemos examinar la otra diagonal, la correlación entre democracia (la puntuación de la democracia) en el momento 2 y paz (la puntuación de la paz) en el momento 1. Esta correlación capta cualquier causalidad inversa, junto con las confusiones que han permanecido a lo largo de la década. Si la primera correlación (causa pasada con efecto presente) es más fuerte que la segunda (efecto pasado con causa presente), ese es un indicio de que la democracia

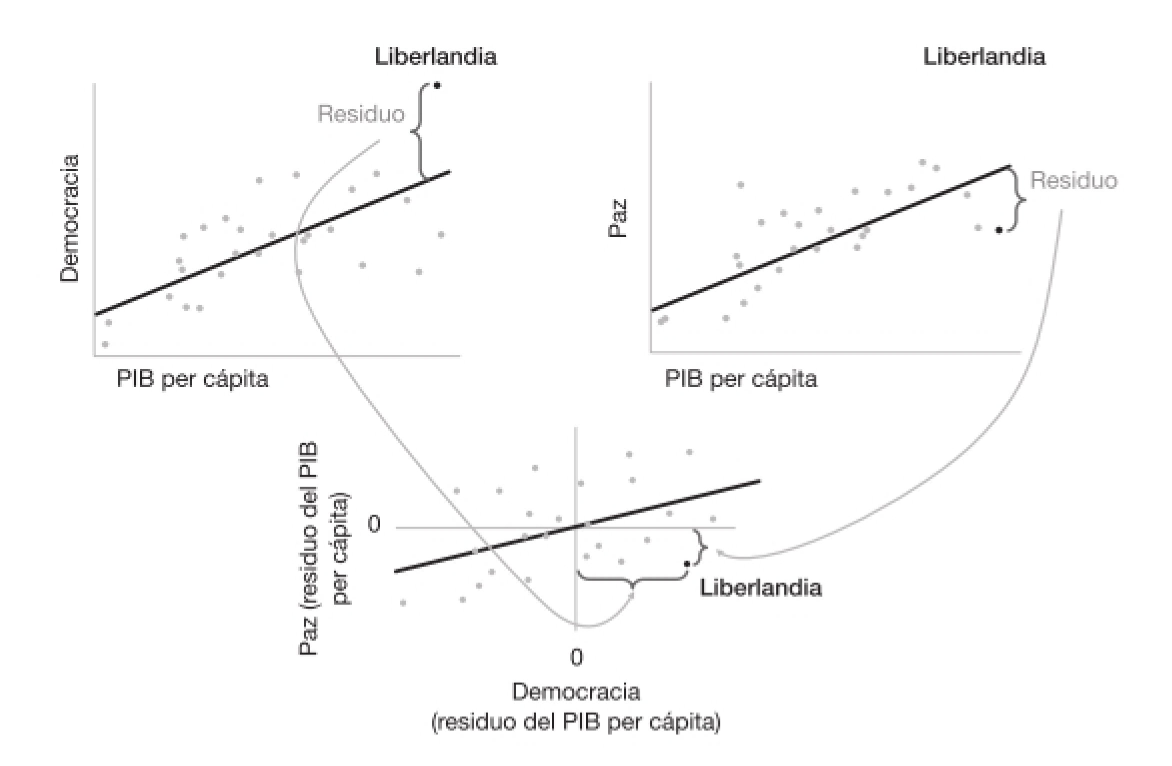
causa la paz más que viceversa. La técnica se denomina *correlación cruzada en panel*, donde el término técnico *panel* designa un conjunto de datos que contiene mediciones en varios puntos temporales.

También las confusiones pueden ser domesticadas mediante estadísticas ingeniosas. Puede que hayas leído en los artículos científicos que los investigadores «mantienen constante» o «controlan estadísticamente» alguna variable de confusión o de estorbo. La forma más simple de hacer esto se denomina *emparejamiento*.[32] La relación entre democracia y paz está infestada de numerosas confusiones, como la prosperidad, la educación, el comercio y la pertenencia a organizaciones creadas en virtud de tratados. Consideremos una de ellas, la prosperidad, medida como producto interior bruto (PIB) per cápita. Supongamos que, para cada democracia de nuestra muestra, encontramos una autocracia que tenga el mismo PIB per cápita. Si comparamos las puntuaciones medias de la paz de las democracias con sus equivalentes en las autocracias, tendremos una estimación de los efectos de la democracia sobre la paz, manteniendo constante el PIB. La lógica del emparejamiento es sencilla, pero requiere una gran reserva de candidatos en la que encontrar buenos emparejamientos, y el número se dispara a medida que han de mantenerse constantes más confusiones. Esto puede funcionar para un estudio epidemiológico con decenas de miles de participantes entre los que elegir, pero no para un estudio político en un mundo con solo 193 países.

La técnica más general se conoce como *regresión múltiple*, y saca provecho del hecho de que una confusión nunca guarda una correlación *perfecta* con una supuesta causa. Las discrepancias entre ellas no resultan ser ruido molesto, sino información reveladora. Veamos cómo podría funcionar esto con la democracia, la paz y el PIB per cápita. En primer lugar, representamos la supuesta causa, la puntuación de la democracia, en relación con la variable estorbo (gráfica superior izquierda), un punto por país (los datos son falsos, inventados para ilustrar la lógica). Trazamos la línea de regresión y dirigimos nuestra atención a los residuos: la distancia vertical entre cada punto y la línea, correspondiente a la discrepancia entre lo democrático que *sería* un país si la renta predijera perfectamente la democracia y lo democrático que es en realidad. Acto seguido nos

deshacemos de la puntuación original de la democracia de cada país y la sustituimos por el residuo: la medida de cuán democrático es, controlando su renta.

A continuación hacemos lo mismo con el supuesto efecto, la paz. Representamos la puntuación de la paz en función de la variable estorbo (gráfica superior derecha), medimos los residuos, nos deshacemos de los datos originales de la paz y los reemplazamos por los residuos, esto es, cuán pacífico es cada país más allá de lo que cabría esperar de su renta. El paso final es evidente: correlacionar los residuos de paz con los residuos de democracia (gráfica inferior). Si la correlación es significativamente diferente de cero, cabe aventurar que la democracia causa la propensión a la paz, manteniendo constante la prosperidad.



Lo que acabamos de ver es el núcleo de la inmensa mayoría de las estadísticas utilizadas en epidemiología y en ciencias sociales, el llamado *modelo lineal general*. El producto final es una ecuación que nos permite predecir el efecto de una suma ponderada de los predictores (algunos de ellos, presumiblemente, causas). Si eres un buen pensador visual, puedes imaginar la predicción como un *plano* inclinado, más que como una línea,

que flota por encima del suelo definido por los dos predictores. Puede agregarse cualquier número de predictores, creando un hiperplano en el hiperespacio; esto desborda rápidamente nuestras débiles capacidades de imaginación visual (que tiene suficientes problemas con tres dimensiones), pero en la ecuación consiste únicamente en añadir más términos a la serie. En el caso de la paz, la ecuación podría ser: paz = (a × democracia) + (b × PIB per cápita) + (c × comercio) + (d × pertenencia a tratados) + (e × educación), suponiendo que cada uno de estos cinco podría ser un estímulo o un obstáculo para la paz. El análisis de regresión nos informa de cuál de las variables candidatas influye en la predicción del resultado, manteniendo constante cada una de las otras. No es una máquina preparada para demostrar la causalidad —todavía tenemos que interpretar las variables y las conexiones plausibles entre estas, y tener cuidado con un sinnúmero de trampas—, pero es la herramienta más frecuentemente utilizada para desenmarañar las múltiples causas y factores de confusión.

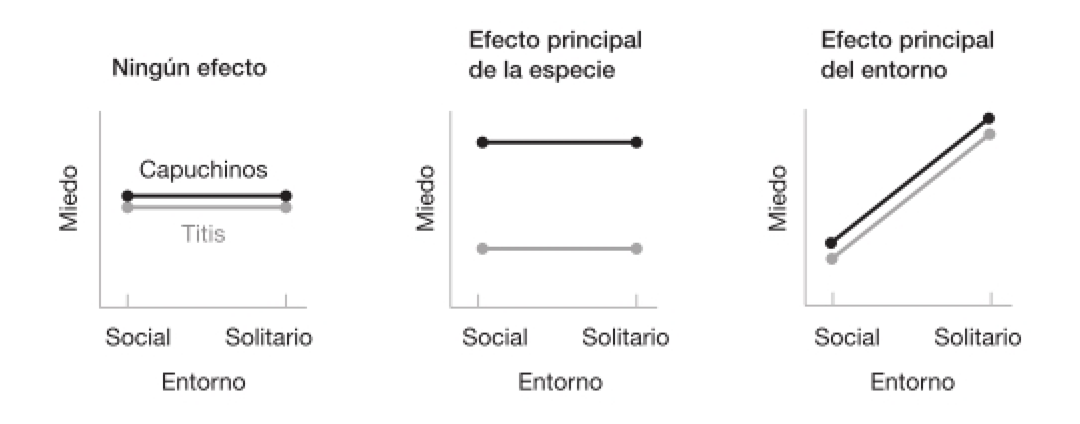
MÚLTIPLES CAUSAS, ADICIONES E INTERACCIONES

El álgebra de una ecuación de regresión es menos importante que la gran idea ostentada por su forma: los sucesos tienen más de una causa, todas ellas estadísticas. La idea parece elemental, pero es desdeñada habitualmente en el discurso público. Con excesiva frecuencia, la gente escribe como si todos los resultados tuvieran una única causa indefectible: si se ha mostrado que A influye en B, eso demuestra que C no puede influir en ella. Los virtuosos dedican diez mil horas a la práctica de su arte u oficio; se dice que esto demuestra que el éxito es una cuestión de práctica, no de talento. Hoy en día los hombres lloran dos veces más que sus padres; esto demuestra que la diferencia en el llanto entre hombres y mujeres es social más que biológica. La posibilidad de múltiples causas (naturaleza *y* crianza, talento *y* práctica) es inconcebible.

Más esquiva aún es la idea de la *interacción* de causas: la posibilidad de que el efecto de una causa dependa de otra. Tal vez todo el mundo se beneficie de la práctica, pero las personas talentosas se benefician más. Lo que necesitamos es un vocabulario para hablar de las causas múltiples y

pensar en ellas. Este es otro ámbito en el que unos pocos conceptos simples de la estadística pueden hacernos más inteligentes a todos. Los conceptos reveladores son *efecto principal* e *interacción*.

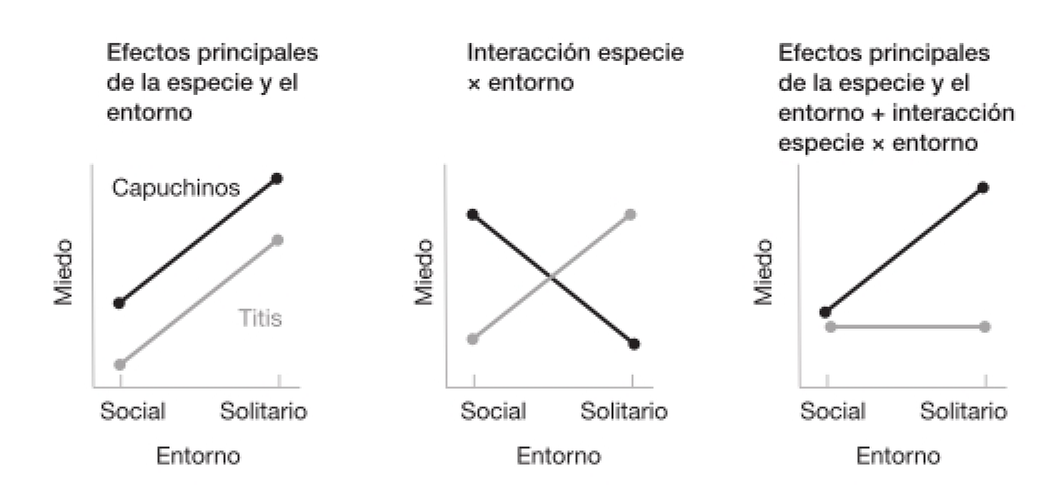
Permítaseme ilustrarlos con datos falsos. Supongamos que estamos interesados en lo que hace temerosos a los monos: la herencia, es decir, la especie a la que pertenecen (capuchinos o titís), o el entorno en el que se han criado (solos con sus madres o en un gran recinto con otras muchas familias de monos). Supongamos que disponemos de una forma de medir el miedo; por ejemplo, cuánto se acerca el mono a una serpiente de goma. Con dos posibles causas y un efecto, pueden suceder seis cosas diferentes. Esto suena complicado, pero las posibilidades saltan a la vista desde la página tan pronto como las representamos en gráficas. Comencemos con las tres más simples.



La gráfica de la izquierda muestra una nada como una casa: un mono es un mono. La especie no importa (las líneas caen una encima de otra); el entorno tampoco importa (cada línea es horizontal). La gráfica del medio es lo que veríamos si la especie importara (los capuchinos son más asustadizos que los titís, lo cual se muestra mediante su línea, que flota más alta en la gráfica), mientras que el entorno no (ambas especies son igual de miedosas si se crían solas o en compañía, como se aprecia porque cada línea es horizontal). En la jerga, decimos que hay un *efecto principal* de la especie, lo cual significa que el efecto se observa en general, con independencia del entorno. La gráfica de la derecha muestra el resultado opuesto, un efecto principal del entorno pero ninguno de la especie. Crecer solo vuelve más

miedoso al mono (como se ve en la pendiente de las líneas), pero lo hace por igual en los capuchinos y en los titís (como se aprecia en el hecho de que las líneas caigan una encima de otra).

Hagámonos ahora más inteligentes aún y tratemos de comprender las múltiples causas. Una vez más tenemos tres posibilidades. ¿Qué ocurriría si tanto la especie como el entorno importasen: si los capuchinos fuesen de manera innata más temerosos que los titís *y* si ser criado en solitario volviese más miedoso a un mono? La gráfica de la izquierda muestra esta situación, a saber, dos efectos principales. Las dos líneas tiene pendientes paralelas y una planea sobre la otra.



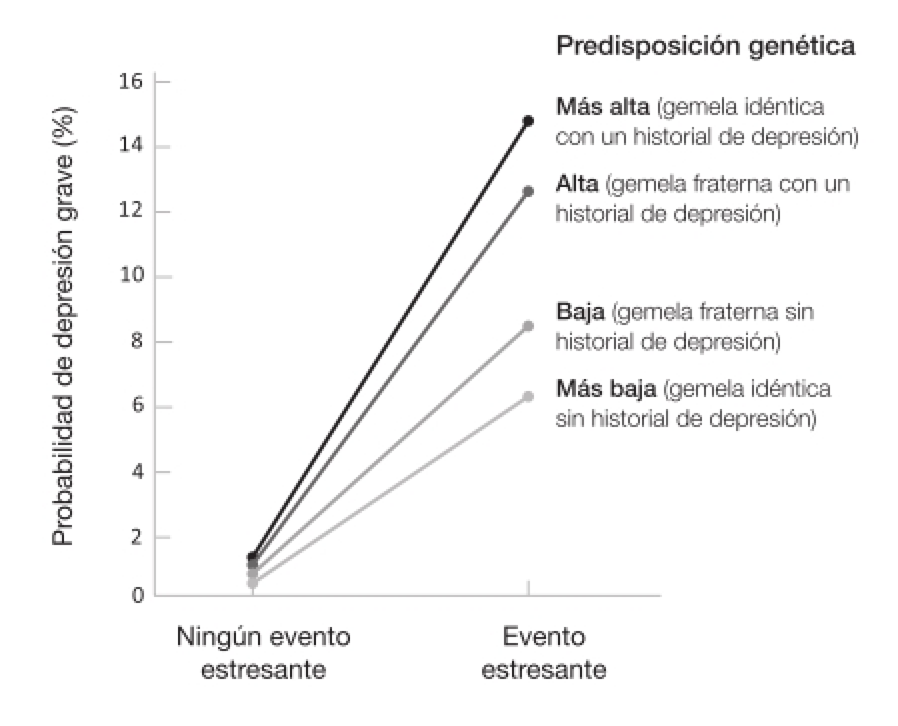
Las cosas se ponen realmente interesantes en la gráfica del medio. En este caso importan ambos factores, pero cada uno depende del otro. Si eres un capuchino, al criarte en solitario te vuelves más valiente; si eres un tití, al criarte en solitario te vuelves más asustadizo. Vemos una *interacción* entre la especie y el entorno, que visualmente se plasma en que las líneas no son paralelas. En estos datos, las líneas se cruzan en una X perfecta, lo cual significa que los efectos principales se contrarrestan totalmente. En general, la especie no importa: el punto medio de la línea de los capuchinos se encuentra encima del punto medio de la línea de los titís. El entorno tampoco importa en general: la media de social, que corresponde al punto medio entre los dos extremos de la izquierda, se alinea con la media de

solitario, correspondiente al punto medio entre los extremos de la derecha. Por supuesto, la especie y el entorno importan; lo que ocurre es que la importancia de cada factor depende del otro.

Finalmente, una interacción puede coexistir con uno o más efectos principales. En la gráfica de la derecha, criarse en solitario hace más miedosos a los capuchinos, pero no ejerce efecto alguno en los siempre tranquilos titís. Dado que el efecto en los titís no contrarresta perfectamente el efecto en los capuchinos, vemos un efecto principal de la especie (la línea de los capuchinos es más alta) y un efecto principal del entorno (el punto medio de los dos extremos izquierdos está más bajo que el punto medio de los dos derechos). Pero cuando interpretamos un fenómeno con dos o más causas, cualquier interacción reemplaza los efectos principales: nos ofrece más información sobre lo que sucede. Una interacción implica habitualmente que ambas causas se entremezclan en un único eslabón de la cadena causal, en vez de producirse en diferentes eslabones y luego simplemente sumarse. Con estos datos, el eslabón común podría ser la amígdala, la parte del cerebro que registra las experiencias de miedo, que puede ser plástica en los capuchinos pero estar programada en los titís.

Con estas herramientas cognitivas, ahora estamos equipados para comprender las múltiples causas en el mundo: podemos ir más allá de «naturaleza versus crianza» y de si los genios «nacen o se hacen». Pasemos a considerar algunos datos reales.

Qué causa la depresión grave, ¿un acontecimiento estresante o una predisposición genética? Esta gráfica representa la probabilidad de sufrir un episodio de depresión severa en una muestra de mujeres con hermanas gemelas.[33]



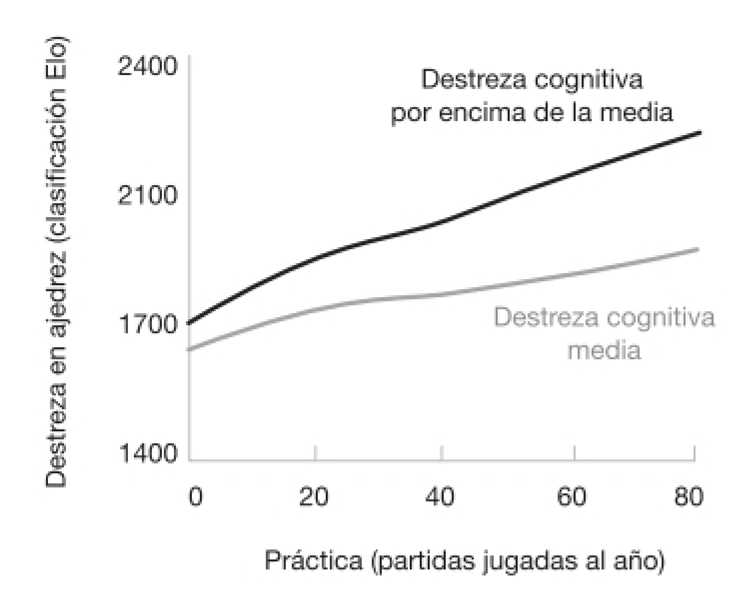
Adaptado con el permiso de Kendler, Kessler, *et al*., 2010

La muestra incluye a mujeres que habían sufrido un estresor severo, como un divorcio, una agresión sexual o la muerte de un pariente cercano (los puntos de la derecha), y a mujeres que no habían sufrido ninguno (los puntos de la izquierda). Al examinar las líneas de arriba abajo, la primera corresponde a las mujeres que pueden tener una alta predisposición a la depresión, porque su gemela idéntica, con la que comparten todos sus genes, la padeció. La línea siguiente corresponde a las mujeres con una cierta predisposición a la depresión, porque un gemelo fraterno o un mellizo, con quien comparten *la mitad* de sus genes, la sufrió. Debajo de esta tenemos una línea para las mujeres que no están particularmente predispuestas, porque su gemelo fraterno no sufrió depresión. Abajo del todo encontramos una línea para las mujeres que tienen el riesgo más bajo, ya que su gemela idéntica no la padeció.

El patrón de la gráfica nos revela tres cosas. La experiencia importa: vemos un efecto principal del estrés en la pendiente ascendente del abanico de líneas, que muestra que sufrir un evento estresante aumenta las probabilidades de deprimirse. En conjunto, los genes importan: las cuatro líneas flotan a diferentes alturas, mostrando que cuanto más alta es la

predisposición genética, mayor es la probabilidad de sufrir un episodio depresivo. Pero la auténtica clave es la *interacción*: las líneas no son paralelas (otra manera de decirlo es que los puntos coinciden en la izquierda, pero están dispersos a la derecha). Si no sufres un evento estresante, tus genes apenas importan: independientemente de tu genoma, la probabilidad de un episodio depresivo es inferior al 1 %. Pero si sufres un evento estresante, tus genes importan mucho: una dosis completa de genes asociados con escapar de la depresión mantiene el riesgo de deprimirse en un 6 % (línea más baja); una dosis completa de genes asociados con sufrir la depresión duplica con creces el riesgo hasta el 14 % (línea más alta). La interacción no solo nos revela que tanto los genes como el entorno son importantes, sino también que ambos parecen ejercer sus efectos en el mismo eslabón de la cadena causal. Los genes que estas gemelas comparten en diferentes grados no son genes de la depresión *per se*; son genes de la vulnerabilidad o la resiliencia ante experiencias estresantes.

Pasemos al asunto de si las estrellas nacen o se hacen. La gráfica de la página siguiente, que también corresponde a un estudio real, muestra las clasificaciones de la destreza en ajedrez en una muestra de jugadores de toda la vida que difieren en la medida de su capacidad cognitiva y en el número de partidas que juegan al año.[34] Muestra que la práctica nos hace mejores, si no perfectos: vemos un efecto principal de las partidas jugadas al año, apreciable en la pendiente ascendente general. El talento cuenta: observamos un efecto principal de la destreza, visible en la diferencia entre las dos líneas. Pero la verdadera moraleja de la historia es su *interacción*: las líneas no son paralelas, lo cual muestra que los jugadores más inteligentes ganan más con cada partida adicional de práctica. Una manera equivalente de expresar esto es que, sin la práctica, la destreza cognitiva apenas importa (los extremos izquierdos de las líneas casi se superponen), pero con la práctica, los jugadores más inteligentes exhiben su talento (los extremos derechos están separados). El conocimiento de la diferencia entre los efectos principales y las interacciones no solo nos protege de caer en falsas dicotomías, sino que nos brinda asimismo una comprensión más profunda de la naturaleza de las causas subyacentes.



Adaptado con el permiso de Vaci, Edelsbrunner, *et al.*, 2019

REDES CAUSALES Y SERES HUMANOS

Como un modo de comprender la riqueza causal del mundo, una ecuación de regresión es muy simple: se limita a sumar un montón de predictores ponderados. También pueden incorporarse las interacciones, que pueden representarse como predictores adicionales derivados de la multiplicación de los predictores interactuantes. Una ecuación de regresión no es ni de lejos tan compleja como las redes de aprendizaje profundo que vimos en el capítulo 3, que incorporan millones de variables y las combinan en largas e intrincadas cadenas de fórmulas, en lugar de limitarse a echarlas a una tolva y sumarlas. No obstante, pese a su simplicidad, uno de los descubrimientos impresionantes de la psicología del siglo XX es que una simple ecuación de regresión suele producir mejores resultados que un experto humano. El hallazgo, observado por vez primera por el psicólogo Paul Meehl, responde al nombre de *juicio clínico versus actuarial*.[35]

Supongamos que queremos predecir algún resultado cuantificable: cuánto sobrevivirá un paciente con cáncer; si un paciente psiquiátrico acaba siendo diagnosticado de neurosis leve o de psicosis severa; si un acusado se fugará estando bajo fianza, se saltará la libertad condicional o reincidirá;

qué resultados obtendrá un estudiante en una escuela de posgrado; si un negocio tendrá éxito o se irá al garete; que rentabilidad tendrá un fondo de acciones. Tenemos un conjunto de predictores: una lista para cotejar síntomas, un conjunto de características demográficas, un recuento de conductas pasadas, un certificado de calificaciones universitarias o resultados de pruebas, o cualquier elemento que podría ser relevante para el reto de la predicción. Le mostramos los datos a un experto (un psiquiatra, un juez, un analista de inversiones, etcétera) y, al mismo tiempo, los introducimos en un análisis de regresión estándar para obtener la ecuación de la predicción. ¿Quién es el pronosticador más preciso, el experto o la ecuación?

La ganadora es, casi siempre, la ecuación. De hecho, un experto a quien se le facilita la ecuación y se le permite que la utilice para complementar su juicio logra con frecuencia peores resultados que la ecuación por sí sola. La razón es que los expertos se apresuran demasiado a ver circunstancias atenuantes que, a su parecer, tornan inaplicable la fórmula en cuestión. Esto se conoce a veces como el *problema de la pierna rota*, en virtud de la idea de que un experto humano, a diferencia de un algoritmo, tiene la capacidad de saber que un tipo que se acaba de romper la pierna no irá a bailar esa noche, incluso si una fórmula predice que lo hace todas las semanas. El problema estriba en que la ecuación *ya* tiene en cuenta la probabilidad de que las circunstancias atenuantes alteren el resultado y las incluye en la mezcla junto con todas las demás influencias, en tanto que el experto humano está demasiado impresionado con los detalles llamativos y tira por la ventana con excesiva rapidez las tasas base. De hecho, los análisis de regresión revelan que algunos de los predictores en los que más confían los expertos humanos, como las entrevistas cara a cara, son perfectamente inútiles.

No se trata de prescindir de los humanos. Una persona continúa siendo indispensable a la hora de suministrar predictores que requieren una comprensión real, como entender el lenguaje y categorizar el comportamiento. Se trata simplemente de que el humano es inepto al *combinarlos*, mientras que esa es la especialidad de un algoritmo de regresión. Como advierte Meehl, en la caja del supermercado no le dirías al

cajero: «Me parece que el total ronda los setenta y seis dólares, ¿es correcto?». Sin embargo, eso es lo que hacemos cuando combinamos intuitivamente un conjunto de causas probabilísticas.

Pese a todo el poder de una ecuación de regresión, el descubrimiento más aleccionador acerca de la predicción del comportamiento humano es lo impredecible que este es. Resulta fácil decir que el comportamiento es causado por una combinación de la herencia y el ambiente. Ahora bien, cuando nos fijamos en un predictor que tiene que ser más potente que la mejor ecuación de regresión (un gemelo idéntico de una persona, que comparta su genoma, familia, vecindario, escolarización y cultura), observamos que la correlación entre las características de ambos gemelos, aunque mucho más alta que el azar, es mucho más baja de uno, habitualmente en torno a 0,6.[36] Esto deja misteriosamente sin explicar *muchas* de las diferencias humanas: pese a las causas cuasi idénticas, los efectos distan de ser idénticos. Un gemelo puede ser homosexual y el otro heterosexual, uno esquizofrénico y el otro funcionar con normalidad. En la gráfica de la depresión, hemos visto que la probabilidad de que una mujer sufra depresión si vive un acontecimiento estresante *y* tiene una disposición genética demostrada a la depresión no es del cien por cien, sino solo del 14 %.

Un impresionante estudio reciente refuerza la idea de la obstinada impredecibilidad de la especie humana.[37] A ciento sesenta equipos de investigadores se les proporcionó un masivo conjunto de datos sobre miles de familias vulnerables, entre los que figuraban sus ingresos, su educación, sus historiales médicos y los resultados de múltiples entrevistas y evaluaciones en el domicilio. Se retó a los equipos a predecir los resultados de las familias en aspectos como las calificaciones de los hijos, o la probabilidad de que los padres fuesen desahuciados, contratados o inscritos para recibir capacitación laboral. A los competidores se les permitía aplicar al problema cualquier algoritmo que desearan: la regresión, el aprendizaje profundo o cualquier otra moda o tendencia en inteligencia artificial. ¿Cuáles fueron los resultados? En las sencillas palabras del resumen del artículo: «Las mejores predicciones no fueron muy exactas». Las características idiosincrásicas de cada familia desbordaban los predictores

genéricos, independientemente del ingenio con el que estos se combinasen. Esto supone una tranquilidad para aquellos a quienes les preocupa que la inteligencia artificial no tarde en predecir cada uno de nuestros movimientos. Pero es asimismo un jarro de agua fría para nuestras pretensiones de comprender cabalmente la red causal en la que nos hallamos insertos.

Y hablando de humildad, hemos llegado al final de siete capítulos destinados a pertrecharte de lo que, a mi juicio, son las herramientas más importantes de la racionalidad. Si lo he conseguido, apreciarás estas últimas palabras de *XKCD*.







xkcd.com

CAPÍTULO

10

¿Qué le pasa a la gente?

Dile a la gente que hay un hombre invisible en el cielo que creó el universo y la inmensa mayoría te creerá. Dile que la pintura está húmeda y tendrá que tocar para asegurarse.

GEORGE CARLIN

Este es el capítulo que la mayoría de vosotros estabais esperando. Lo sé por las conversaciones y la correspondencia. En cuanto menciono el tema de la racionalidad, la gente me pregunta por qué la humanidad parece estar perdiendo la cabeza.

Mientras escribo estas páginas, se vislumbra un glorioso hito en la historia de la racionalidad: las vacunas que probablemente acaben con una plaga mortal se están administrando menos de un año después de la aparición de la plaga. Sin embargo, en ese mismo año, la pandemia de la COVID-19 desató un carnaval de disparatadas teorías de la conspiración: que la enfermedad era un arma biológica diseñada en un laboratorio chino, un bulo propagado por el propio Partido Demócrata para sabotear las probabilidades de reelección de Donald Trump, un subterfugio de Bill Gates para implantar microchips rastreables en los cuerpos de los individuos, un complot de una camarilla de élites globales para controlar la economía mundial, un síntoma del lanzamiento de las redes de datos móviles de quinta generación y un medio para que Anthony Fauci (director del Instituto Nacional de Alergias y Enfermedades Infecciosas) obtuviese unos beneficios extraordinarios con una vacuna.[1] Poco antes del anuncio de las vacunas, un tercio de los estadounidenses decían que las rechazarían, como parte de un movimiento antivacunas que se opone a la invención más benéfica de la historia de nuestra especie.[2] La charlatanería en torno a la COVID ha recibido el respaldo de celebridades, políticos y, de manera inquietante, de la persona más poderosa de la Tierra en el momento de la pandemia, el presidente de Estados Unidos Donald Trump.

El propio Trump, sistemáticamente apoyado por cerca del 40 % de los estadounidenses, suscitó otras dudas a lo largo de su presidencia sobre nuestra capacidad colectiva para razonar. En febrero de 2020, predijo que la COVID-19 desaparecería «como un milagro» y respaldaba remedios de curanderos como medicamentos para la malaria, inyecciones de lejía y sondas de luz. Desdeñaba medidas básicas de salud pública como las mascarillas y la distancia, incluso después de contraer él mismo el virus, inspirando a millones de estadounidenses a incumplir las medidas, y

aumentando el número de muertes y las dificultades económicas.[3] Todo ello formaba parte de un rechazo de mayor envergadura de las reglas de la razón y de la ciencia. Trump dijo alrededor de treinta mil mentiras durante su mandato, tuvo un secretario de Prensa que pregonaba «hechos alternativos», afirmaba que el cambio climático era un bulo chino y ocultaba los conocimientos de los científicos de las agencias federales que supervisaban la salud pública y la protección ambiental.[4] Publicitó reiteradamente a QAnon, el culto a la conspiración que cuenta con millones de seguidores y que le atribuye el mérito de combatir a una camarilla de pedófilos adoradores de Satán inserta en el «Estado profundo» estadounidense. Y se negó a reconocer su derrota en las elecciones de 2020, librando disparatadas batallas legales para anular los resultados, dirigidas por abogados que citaban otra conspiración, en esta ocasión de Cuba, Venezuela y varios gobernadores y funcionarios de su propio partido.

La charlatanería sobre la COVID, la negación del cambio climático y las teorías de la conspiración son síntomas de lo que algunos están definiendo como *crisis epistemológica* y *era de la posverdad*.[5] Otro síntoma son las *fake news* o noticias falsas. En la segunda década del siglo XXI, los medios sociales se han convertido en canales para torrentes de historias fantásticas como estas:[6]

- «El papa Francisco conmociona al mundo, respalda a Donald Trump como presidente».
- «Yoko Ono: “Tuve una aventura con Hillary Clinton en los años setenta”.»
- «Los demócratas votan ahora para mejorar la atención médica a los ilegales, votan contra los veteranos que llevan diez años esperando por el mismo servicio.»
- «Trump prohíbe todos los programas de televisión que promuevan actividades de los gais.»
- «Una mujer presenta una demanda contra Samsung por valor de 1,8 millones de dólares al quedarse atascado su móvil en su vagina.»
- «Ganador de lotería arrestado por volcar doscientos mil dólares de estiércol en el césped de su exjefe.»

También proliferan las creencias en los espíritus malignos, la magia negra y otras supersticiones. Como mencioné en el primer capítulo, tres cuartas partes de los estadounidenses profesan al menos una creencia paranormal. He aquí algunas cifras de la primera década de nuestro siglo:[7]

• Posesión por el demonio, 42 %.
• Percepción extrasensorial, 41 %.
• Fantasmas y espíritus, 32 %.
• Astrología, 25 %.
• Brujas, 21 %.
• Comunicación con los muertos, 29 %.
• Reencarnación, 24 %.
• Energía espiritual en montañas, árboles y cristales, 26 %.
• Mal de ojo, maldiciones, hechizos, 16 %.
• Consultas a adivinos o videntes, 15 %.

Igual de inquietante para alguien como yo, que gusta de trazar el progreso humano, resulta el hecho de que estas creencias muestren pocos signos de disminuir a lo largo de las décadas, y de que las generaciones más jóvenes no sean más escépticas que sus mayores (en lo que atañe a la astrología son más crédulas).[8]

También es popular una miscelánea de embustes que el historiador de la ciencia Michael Shermer denomina *creencias extrañas*.[9] Mucha gente apoya teorías conspiratorias como la negación del Holocausto, los complots para el asesinato de Kennedy y la teoría sobre la «verdad oculta» del 11S, según la cual las Torres Gemelas fueron derribadas mediante una demolición controlada para justificar la invasión estadounidense de Irak. Varios videntes, cultos e ideologías han convencido a sus seguidores de que el fin del mundo se acerca; discrepan sobre el cuándo, pero se apresuran a posfechar sus predicciones cuando se ven desagradablemente sorprendidos al descubrirse viviendo otro día. Y entre una cuarta parte y un tercio de los estadounidenses creen que nos han visitado los extraterrestres, ya sean los

contemporáneos, que mutilan al ganado y fecundan a las mujeres para engendrar híbridos de alienígenas y humanos, ya sean los antiguos, que construyeron las pirámides y las estatuas de la isla de Pascua.

¿CÓMO PODEMOS EXPLICAR esta pandemia de necedades? Como a Charlie Brown en la tira cómica de *Peanuts*, puede provocar dolor de estómago, especialmente cuando aparece Lucy en representación de una gran porción de nuestros compatriotas:



Comencemos por dejar a un lado tres explicaciones populares, no porque sean falsas, sino porque son demasiado simplistas para resultar satisfactorias. He de admitir que la primera de ellas es el inventario de las falacias lógicas y estadísticas explicadas en los capítulos precedentes. Sin lugar a dudas, muchas supersticiones se originan en la sobreinterpretación de las coincidencias, en la incapacidad de calibrar las evidencias en función de las probabilidades previas, en la generalización excesiva a partir de las anécdotas y en el salto de la correlación a la causalidad. Un excelente ejemplo es la idea errónea de que las vacunas causan autismo, reforzada por la observación de que los síntomas autistas aparecen casualmente en torno a la edad en la que se vacuna por primera vez a los niños. Y todas ellas representan fallos del pensamiento crítico y de la fundamentación de la creencia en las evidencias; eso es lo que nos da derecho a decir que son falsas de entrada. Sin embargo, en el laboratorio de psicología cognitiva nada podría haber predicho QAnon, ni es probable que sus seguidores se desengañen gracias a un tutorial de lógica o de probabilidad.

Una segunda iniciativa poco prometedora consiste en culpar de la irracionalidad actual al chivo expiatorio para todo en nuestros días: los medios sociales. Las teorías de la conspiración y las falsedades virales son probablemente tan viejas como el lenguaje.[10] Al fin y a la postre, ¿qué son los relatos de milagros en las Escrituras sino noticias falsas acerca de fenómenos paranormales? Durante siglos, los judíos han sido acusados de conspirar para envenenar pozos, sacrificar niños cristianos, controlar la economía mundial y fomentar los levantamientos comunistas. En muchos momentos de la historia, a otras razas, minorías y gremios se les han atribuido asimismo nefarios complots y han sido blancos de la violencia.[11] Los politólogos Joseph Uscinski y Joseph Parent rastrearon la popularidad de las teorías conspiratorias en cartas al editor de importantes periódicos estadounidenses desde 1890 hasta 2010 y no hallaron cambio alguno durante ese periodo; tampoco aumentaron las cifras en la década subsiguiente.[12] En cuanto a las *fake news*, antes de que se difundieran por Twitter o Facebook, extravagantes episodios que le ocurrían a un amigo de un amigo circulaban como leyendas urbanas (la niñera *hippie*, la rata frita de Kentucky, los sádicos de Halloween) o eran celebrados en las portadas de tabloides de supermercado («Bebé nace hablando: describe el cielo»; «Dick Cheney es un robot»; «Unos cirujanos trasplantan la cabeza de un niño al cuerpo de su hermana»).[13] Los medios sociales pueden estar acelerando en efecto su propagación, pero el apetito de floridas fantasías mora en lo más profundo de la naturaleza humana: son las personas, no los algoritmos, las que crean estas historias, y es a ellas a quienes apelan. Y pese a todo el pánico sembrado por las noticias falsas, su impacto político es leve: despiertan el interés de una facción de partidarios más que influir en una masa de indecisos.[14]

Finalmente, hemos de ir más allá de las excusas improvisadas que se limitan a atribuir la irracionalidad a otros. Jamás es una explicación satisfactoria decir que las personas abrazan una creencia falsa porque esta las consuela o las ayuda a dar sentido al mundo, porque eso no hace sino plantear la pregunta de *por qué* los individuos habrían de hallar consuelo y sentido de finalidad en creencias que no podrían hacerles ningún bien. La realidad ejerce una poderosa presión selectiva. Un homínido que se

confortara creyendo que un león era una tortuga o que comer arena alimentaría su cuerpo tendría más dificultades para reproducirse que sus rivales basados en la realidad.

Tampoco basta con tachar a los humanos de irremediablemente irracionales. Al igual que nuestros antepasados cazadores recolectores vivían de su ingenio en ecosistemas implacables, hoy en día los teóricos de la conspiración y los creyentes en los milagros superan las exigentes pruebas de sus propios mundos: mantienen trabajos, educan a sus hijos y tienen un techo sobre su cabeza y comida en el frigorífico. Es más, una réplica favorita de los defensores de Trump a la acusación de que este padecía un deterioro cognitivo era: «Si es tan estúpido, ¿cómo llegó a ser presidente?». Y, a menos que creamos que los científicos y los filósofos son una raza superior de humanos, hemos de reconocer que la mayoría de los miembros de nuestra especie poseen la capacidad para descubrir y aceptar los cánones de la racionalidad. A fin de comprender los delirios populares y la locura de las multitudes, tenemos que examinar las facultades cognitivas que funcionan bien en algunos entornos y para algunos propósitos, pero fracasan cuando se aplican a gran escala, en circunstancias novedosas o al servicio de otras metas.

RAZONAMIENTO MOTIVADO

La racionalidad es desinteresada. Es la misma para todo el mundo en todas partes, con un rumbo y un impulso propios. Por ese motivo la racionalidad puede ser una molestia, un impedimento y una afrenta. En la novela de Rebecca Newberger Goldstein *36 Arguments for the Existence of God: A Work of Fiction [Treinta y seis argumentos a favor de la existencia de Dios: una obra de ficción]*, un eminente estudioso de la literatura explica a un estudiante de posgrado por qué aborrece el pensamiento deductivo:[15]

> Es una forma de tortura para las personas imaginativas, el totalitarismo mismo del pensamiento, en el que una línea sigue estrictamente a la otra y todas conducen inexorablemente a una única conclusión constante. Una prueba de Euclides me recuerda más que nada a las tropas que marchaban a paso de ganso ante el Dictador Supremo. Siempre me he deleitado en la negativa de mi mente a seguir una sola línea de cualquier explicación matemática que se me ofreciese. ¿Por qué habrían de obtener algo de mí estas ciencias exactas? O como sostiene

sagazmente el hombre del subsuelo de Dostoievski: «Buen Dios, ¿qué me importan las leyes de la naturaleza y la aritmética si, por una u otra razón, no me gustan estas leyes, incluida esa que dice "dos y dos son cuatro"?». Dostoievski desdeñaba la lógica hegemaniaca y yo no puedo hacer menos.

El motivo evidente por el que los individuos evitan montar en un tren de razonamientos es que no les gusta adónde los lleva. Puede terminar en una conclusión que no sea de su interés, como una asignación de dinero, poder o prestigio que sea objetivamente justa, pero que beneficia a otros. Como señalaba Upton Sinclair: «Resulta difícil conseguir que un hombre entienda algo cuando su salario depende de que no lo entienda».[16]

El método tradicional para desviar una línea de razonamiento antes de que llegue a un destino no deseado consiste en hacer descarrilar al razonador mediante la fuerza bruta. Pero existen métodos menos burdos que explotan las inevitables incertidumbres que rodean cualquier asunto y dirigen el argumento en la dirección favorita con sofistería, manipulación informativa y demás artes de la persuasión. Los dos miembros de una pareja en busca de apartamento, por ejemplo, pueden enfatizar las razones por las que el piso que resulta estar más cerca de donde él o ella trabaja es objetivamente mejor para ambos, como su espacio o su precio asequible. Esta es la materia prima de los argumentos cotidianos.

El acopio de recursos retóricos para conducir un argumento hacia una conclusión predilecta se denomina *razonamiento motivado*.[17] El motivo puede ser llegar a una conclusión agradable, pero puede ser también hacer alarde de la sabiduría, los conocimientos o las virtudes del argumentador. Todos conocemos al fanfarrón de taberna, al campeón de debate, al abogado astuto, al hombre que explica con condescendencia a una mujer, al que compite por llegar más lejos al orinar, al pugilista intelectual que prefiere *tener razón* que *hacer bien* las cosas.[18]

Muchos de los sesgos que pueblan las listas de enfermedades cognitivas son tácticas de razonamiento motivado. En el capítulo 1 vimos el sesgo de confirmación, como en el caso de la tarea de selección, donde las personas a las que se les pide que den la vuelta a las tarjetas que prueban una regla «si P entonces Q», eligen la tarjeta P, que puede confirmarla, pero no la tarjeta «NO Q», que puede refutarla.[19] Resultan ser más lógicas cuando

quieren que la regla sea falsa. Cuando la regla dice que, si alguien tiene un determinado perfil emocional, está en peligro de morir joven, la persona que lo posee comprueba correctamente la regla (y, al mismo tiempo, se tranquiliza), centrándose en las personas que tienen su perfil o en aquellas que vivieron hasta una edad avanzada.[20]

También estamos motivados para regular nuestra dieta informativa. En la asimilación sesgada (o exposición selectiva), los individuos buscan argumentos que ratifiquen sus creencias y se protegen de aquellos que podrían refutarlas[21] (¿quién de nosotros no se regodea leyendo editoriales políticamente agradables, y no se irrita al leer los del otro bando?). Nuestra autoprotección continúa con los argumentos que sintonizan con nosotros. En la evaluación sesgada, desplegamos nuestro ingenio para apoyar los argumentos que respaldan nuestra posición y somos quisquillosos con aquellos que la refutan. Y están las falacias informales clásicas que vimos en el capítulo 3: *ad hominem*, de autoridad, la de subirse al carro, la genética, la afectiva, la del hombre de paja, y la de suma y sigue. Incluso estamos sesgados con respecto a nuestros sesgos. La psicóloga Emily Pronin ha descubierto que, como en la mítica localidad en la que todos los niños están por encima de la media, una gran mayoría de estadounidenses se consideran menos susceptibles a los sesgos cognitivos que el estadounidense medio, y prácticamente nadie se considera más sesgado.[22]

Tantos de nuestros razonamientos parecen hechos a medida para vencer en las discusiones que algunos científicos cognitivos, como Hugo Mercier y Dan Sperber, creen que tal es la función adaptativa del razonamiento.[23] No hemos evolucionado como científicos intuitivos, sino como intuitivos abogados. Aunque los individuos intentan a menudo salirse con la suya con argumentos débiles a favor de sus propias posiciones, se apresuran a detectar falacias en los argumentos ajenos. Por fortuna, esta hipocresía puede movilizarse para hacernos más racionales colectivamente de lo que cualquiera de nosotros lo es a título individual. La ocurrencia que circula entre los veteranos de los comités según la cual el CI de un grupo es igual al CI más bajo de cualquier miembro del grupo dividido entre el tamaño de grupo resulta totalmente falsa.[24] Cuando las personas evalúan una idea en pequeños grupos con la química adecuada, que consiste en que

no están de acuerdo en todo, pero tienen un interés común en hallar la verdad, detectan sus respectivas falacias y puntos ciegos, y suele vencer la verdad. Cuando se propone a los individuos la tarea de selección de Wason, por ejemplo, solo uno de cada diez escoge las tarjetas correctas, pero cuando se distribuyen en grupos, en torno a siete de cada diez aciertan. Todo lo que hace falta es que un miembro vea la respuesta correcta, y casi siempre esa persona convence a las demás.

El sesgo de mi lado

El deseo de los individuos de salirse con la suya o de actuar como sabelotodo puede explicar solamente parte de nuestra irracionalidad pública. Podemos apreciar otra parte considerando este problema en las políticas basadas en las evidencias. Las medidas de control de armas, ¿disminuyen la criminalidad porque son menos los criminales que pueden conseguir una, o bien la incrementan porque los ciudadanos observantes de la ley ya no pueden protegerse?

He aquí los datos de un hipotético estudio que dividía las ciudades en aquellas que adoptaban la prohibición de llevar armas ocultas (primera fila) y aquellas que no lo hacían (segunda fila).[25] En cada columna se indica el número de esas ciudades que vieron mejorar sus índices de criminalidad (columna izquierda) o empeorarlos (columna derecha). A partir de estos datos, ¿concluiríamos que el control de armas es efectivo para reducir la criminalidad?

	Disminución de la tasa de criminalidad	**Aumento de la tasa de criminalidad**
Control de armas	223	75
Sin control de armas	107	21

De hecho, los datos (que son falsos) sugieren que el control de armas *aumenta* la criminalidad. Resulta fácil equivocarse, porque salta a la vista el gran número de ciudades con control de armas en las que descendió la tasa de criminalidad, doscientas veintitrés. Pero eso podría significar únicamente que la criminalidad disminuyó en todo el país, con la medida o sin ella, y que fueron más las ciudades que probaron el control de armas que las que no lo hicieron, una tendencia en las modas políticas. Necesitamos examinar las proporciones. En las ciudades con control de armas, la ratio es aproximadamente de tres a uno (doscientas veintitrés frente a setenta y cinco); en las ciudades sin control, es aproximadamente de *cinco* a uno (ciento siete frente a veintiuna). Por término medio, dicen los datos, en una ciudad iban mejor las cosas sin control de armas que con él.

Como sucede con el test de reflexión cognitiva (capítulo 1), llegar a la respuesta requiere un poco de aritmética: la capacidad de dejar de lado las primeras impresiones y hacer cálculos. Las personas con escasas habilidades numéricas tienden a distraerse con el número grande y concluyen que el control de armas funciona. Pero la auténtica clave de esta ilustración, diseñada por el experto legal Dan Kahan y sus colaboradores, reside en lo que sucedía con los encuestados con conocimientos básicos de aritmética. Los republicanos con competencia matemática tendían a señalar la respuesta correcta, los demócratas con competencia matemática tendían a equivocarse. La razón es que los demócratas *parten* de la creencia de que el control de armas es efectivo y se precipitan a aceptar los datos que muestran que estaban en lo cierto. Los republicanos no pueden soportar la idea y escudriñan los datos con una mirada penetrante que, si son aritméticamente competentes, detecta el patrón real.

Los republicanos podrían atribuir su éxito al hecho de ser más objetivos que los liberales con el corazón sangrante, pero por supuesto los investigadores diseñaron una situación en la que la respuesta incorrecta instintiva era del agrado de los republicanos. Simplemente cambiaron las etiquetas de las columnas, de suerte que los datos sugerían ahora que el control de armas funcionaba: evitaba que se quintuplicase la criminalidad, que solo se triplicaba. En esta ocasión los republicanos con competencia aritmética se ganaban las orejas de burro, mientras los demócratas eran los

Einstein. En una condición de control, el equipo escogía un tema que no despertaba las pasiones ni de los demócratas ni de los republicanos: si una crema solar era o no efectiva para tratar un sarpullido. Sin que ninguna de las facciones se jugara nada en el asunto, tanto los republicanos como los demócratas con conocimientos básicos de aritmética obtenían resultados similares. Un reciente metaanálisis de cincuenta estudios realizado por el psicólogo Peter Ditto y sus colegas confirma el patrón. En un estudio tras otro, liberales y conservadores aceptan o rechazan la misma conclusión científica dependiendo de si esta respalda o no sus temas de discusión, y apoyan o se oponen a la misma medida dependiendo de que esta fuese propuesta por un político demócrata o republicano.[26]

La competencia aritmética políticamente motivada y otras formas de evaluación sesgada muestran que los individuos razonan para conseguir llegar a una conclusión o huir de ella, incluso cuando ello no les ofrezca ninguna ventaja personal. Basta con que la conclusión realce la corrección o la nobleza de su tribu política, religiosa, étnica o cultural. Esto se llama, evidentemente, *sesgo de mi lado*, y se apropia de toda clase de razonamiento, incluida la lógica.[27] Recordemos que la validez de un silogismo depende de su forma, no de su contenido, pero la gente deja que se filtren sus conocimientos y considera válido un argumento si este termina en una conclusión que sabe que es verdadera o desea que lo sea. Lo mismo sucede cuando la conclusión es políticamente atractiva:

- Si las admisiones universitarias son justas, entonces ya no son necesarias las leyes de discriminación positiva.
  Las admisiones universitarias no son justas.
  Por tanto, son necesarias las leyes de discriminación positiva.
- Si las penas menos severas disuaden a las personas de cometer crímenes, entonces no debería utilizarse la pena capital.
  Las penas menos severas no disuaden a las personas de cometer crímenes.
  Por tanto, debería utilizarse la pena capital.

Cuando se pide a los individuos que verifiquen la lógica de estos argumentos, que en ambos casos cometen la falacia de la negación del antecedente, los liberales ratifican erróneamente el primero y rechazan el segundo; los conservadores hacen lo contrario.[28]

En *Duck Soup (Sopa de ganso)*, Chico Marx preguntaba: «¿A quién vas a creer, a mí o a tus propios ojos?». Cuando las personas sucumben al sesgo de mi lado, la respuesta puede que no sea «a mis propios ojos». En una actualización de un estudio clásico que demuestra que los aficionados al fútbol siempre ven más infracciones del equipo rival, Kahan y sus colaboradores mostraban un vídeo de una protesta delante de un edificio.[29] Cuando el título lo etiquetaba como una protesta contra el aborto en un centro de salud, los conservadores veían una manifestación pacífica, mientras que los liberales veían que los manifestantes bloqueaban la entrada e intimidaban a quienes acudían. Cuando se etiquetaba como una protesta contra la exclusión de los homosexuales en un centro de reclutamiento militar, eran los conservadores los que veían horcas y antorchas, y los liberales quienes veían a Mahatma Gandhi.

Una revista refería el estudio sobre el control de armas con el titular «El descubrimiento más deprimente de la historia acerca del cerebro». Ciertamente, existen razones para estar deprimido. Una de ellas es que las opiniones contrarias al consenso científico, como el creacionismo y la negación del cambio climático provocado por los humanos, puede que no sean síntomas de incompetencia aritmética o de analfabetismo científico. Kahan ha descubierto que la mayoría de los creyentes y de los negacionistas están igual de despistados en lo que atañe a los hechos científicos (muchos creyentes en el cambio climático, por ejemplo, piensan que este tiene algo que ver con los vertederos de residuos tóxicos y el agujero de ozono). Lo que predice su creencia es su orientación política: cuanto más a la derecha, más negación.[30]

Otra razón para la tristeza es que, pese a todo lo que se habla de la crisis de la replicabilidad, el sesgo de mi lado es perfectamente replicable. En *The Bias That Divides Us [El sesgo que nos divide]*, el psicólogo Keith Stanovich lo encuentra en todas las razas, los géneros, los estilos cognitivos, los niveles educativos y cuantiles del CI, incluso entre personas

que son demasiado inteligentes como para caer en otros sesgos cognitivos, como el olvido de la tasa base y la falacia del jugador.[31] El sesgo de mi lado no es un rasgo de personalidad general, sino que aprieta cualquier gatillo o tema candente conectado con la identidad del razonador. Stanovich lo relaciona con nuestro momento político. Sugiere que no estamos viviendo en una sociedad de la *posverdad*. El problema estriba en que estamos viviendo en una sociedad *de mi lado*. Los lados son la izquierda y la derecha, y ambos lados creen en la verdad, pero tienen ideas inconmensurables de lo que esta es. El sesgo ha invadido progresivamente nuestras deliberaciones. El espectáculo de las mascarillas durante una pandemia respiratoria convertidas en símbolos políticos es tan solo el síntoma más reciente de la polarización.

SABEMOS DESDE HACE MUCHO TIEMPO que los humanos están muy interesados en dividirse en equipos competitivos, pero no está claro por qué es hoy la división entre izquierda y derecha la que está tirando de la racionalidad de cada lado en direcciones diferentes, en lugar de las líneas de falla habituales de la religión, la raza y la clase. El eje derecha-izquierda se alinea con varias dimensiones morales e ideológicas: jerárquico frente a igualitario, libertario frente a comunitarista, trono y altar frente a Ilustración, tribal frente a cosmopolita, visiones trágicas frente a visiones utópicas, culturas del honor frente a culturas de la dignidad, morales vinculantes frente a morales individualizadoras.[32] Pero recientes golpes de timón en qué apoya cada causa, como la inmigración, el comercio o la simpatía por Rusia, sugieren que los bandos políticos se han convertido en tribus socioculturales más que en ideologías coherentes.

En un diagnóstico reciente, un equipo de científicos sociales concluía que los bandos se asemejan menos a las tribus en sentido literal, que mantienen su unidad mediante el parentesco, que a las sectas religiosas, que se mantienen unidas mediante la fe en su superioridad moral y su desprecio hacia las sectas opuestas.[33] Del ascenso del sectarismo político en Estados Unidos se culpa habitualmente (como de todo lo demás) a los medios sociales, pero sus raíces son más profundas. Estas incluyen el fraccionamiento y la polarización de los medios audiovisuales, toda vez que

la radio hablada y las noticias por cable partidistas desplazan a las cadenas nacionales; los fraudes electorales y otras distorsiones geográficas de la representación política, que incentivan a los políticos a atender a las camarillas más que a las coaliciones; la dependencia de los políticos y de los laboratorios de ideas de donantes ideológicamente comprometidos; la autosegregación de los profesionales liberales educados en enclaves urbanos; y el declive de organizaciones interclasistas de la sociedad civil, tales como iglesias, clubes de servicio y grupos de voluntariado.[34]

¿Podría ser racional el sesgo de mi lado? Existe un argumento bayesiano que dice que uno *debería* sopesar las nuevas evidencias con respecto a la totalidad de sus creencias previas en lugar de creer a pies juntillas cada nuevo estudio. Si el liberalismo ha demostrado ser correcto, entonces uno no debería permitir que un estudio que parezca respaldar una posición conservadora trastoque sus creencias. No es de extrañar que esta fuese la respuesta de varios académicos liberales al metaanálisis de Ditto que sugería que el sesgo político es bipartidista.[35] Nada garantiza que las posiciones favoritas de la izquierda y de la derecha en cualquier momento histórico se alineen con la verdad al 50 %. Incluso si ambos bandos interpretan la realidad a través de sus respectivas creencias, el bando cuyas creencias estén justificadas actuará racionalmente. Tal vez, prosiguen, el bien documentado desequilibrio del mundo académico entre izquierda y derecha no sea un sesgo irracional, sino una calibración acertada de sus probabilidades previas bayesianas al hecho de que la izquierda siempre tiene razón.

La respuesta de los conservadores es (citando a *Hamlet*): «No apliquéis al alma esa unción halagüeña».[36*] Aunque puede que sea cierto que las posiciones izquierdistas estén justificadas con más frecuencia que las de derechas (especialmente si, por la razón que sea, la izquierda congenia más con la ciencia que la derecha), en ausencia de criterios desinteresados ninguno de ambos bandos se encuentra en condiciones de decirlo. Ciertamente, en la historia no escasean los ejemplos de equivocaciones por ambas partes, incluidas algunas realmente extraordinarias.[37] Stanovich advierte que el problema a la hora de justificar

el razonamiento motivado con probabilidades previas bayesianas estriba en que la probabilidad previa refleja con frecuencia lo que el razonador *desea* que sea verdadero, más que lo que tiene *motivos para creer* que lo es.

Existe una racionalidad diferente y más perversa en el sesgo de mi lado, que no dimana de la regla de Bayes, sino de la teoría de juegos. Kahan la denomina *racionalidad expresiva*: el razonamiento impulsado por el objetivo de ser valorados por nuestro grupo de pares, más que de alcanzar la comprensión más precisa del mundo. Las personas expresan opiniones que anuncian dónde mora su corazón. En lo que concierne al destino de quien las expresa en un entorno social, alardear de esas insignias de lealtad no tiene nada de irracional. Expresar una herejía local, como puede ser el rechazo del control de armas en un círculo socialdemócrata o su defensa en uno republicano, puede señalarte como un traidor, un colaboracionista, alguien que «no entiende», y condenarte a la muerte social. De hecho, las mejores creencias indicadoras de la identidad son a menudo las más estrafalarias. Cualquier amigo en las buenas puede decir que la Tierra es redonda, pero solo un hermano de sangre diría que la Tierra es plana, exponiéndose de buen grado al ridículo ante las personas ajenas.[38]

Desgraciadamente, lo que es racional para cada uno de los que buscamos la aceptación en una camarilla no es tan racional para todos nosotros en una democracia que busca la mejor comprensión del mundo. Nuestro problema es que estamos atrapados en una *tragedia de los comunes de la racionalidad*.[39]

DOS CLASES DE CREENCIA: REALIDAD Y MITOLOGÍA

El humor en la tira cómica de *Peanuts*, en la que Lucy queda enterrada en la nieve mientras insiste en que esta brota del suelo, expone una limitación en cualquier explicación de la irracionalidad humana que invoque los motivos ulteriores del razonamiento motivado. Por muy eficazmente que una falsa creencia haga alarde de destreza mental o lealtad a la tribu del creyente, continúa siendo falsa y debería ser castigada por los hechos fríos y duros del mundo. Como escribió el novelista Philip K. Dick,

la realidad es aquello que, aun cuando dejes de creer en ello, no desaparece. ¿Por qué la realidad no obliga a retroceder e impide que la gente crea absurdidades o recompense a aquellos que las afirman y las comparten?

La respuesta es que depende de lo que queramos decir con *creer*. Mercier observa que quienes profesan creencias extrañas a menudo carecen del valor de sus convicciones.[40] Aunque millones de personas respaldaron el rumor de que Hillary Clinton dirigía una red de pedofilia desde el sótano de la pizzería Comet Ping Pong de Washington (la teoría conspiratoria Pizzagate, predecesora de QAnon), prácticamente nadie tomó medidas proporcionales a tamaña atrocidad, como llamar a la policía. La respuesta justa de uno de ellos fue dejar una reseña de una estrella en Google («la *pizza* estaba increíblemente poco hecha. Hombres sospechosos con atuendo profesional en la zona del bar que parecían asiduos no dejaban de mirar a mi hijo y a otros niños del local»). Difícilmente sería esa la reacción de la mayoría de nosotros si pensáramos literalmente que se estaba violando a niños en el sótano. Al menos Edgar Welch, el hombre que irrumpió en la pizzería empuñando su arma en un intento heroico de rescatar a los niños, se tomaba en serio sus creencias. Los otros millones debieron de haber creído el rumor en un sentido muy diferente de *creer*.

Mercier señala asimismo que los creyentes apasionados en enormes y nefarias conspiraciones, como los miembros del movimiento por la verdad del 11S (los *truthers*) y los teóricos de las estelas químicas o *chemtrails* (que sostienen que las estelas químicas de vapor de agua dejadas por los aviones son productos químicos distribuidos en un programa gubernamental secreto para envenenar a la población), publican sus manifiestos y celebran sus reuniones al descubierto, pese a su creencia en un complot brutalmente eficaz por un régimen omnipotente para reprimir a quienes, como ellos, dicen la verdad con valentía. No es esa la estrategia que vemos en los disidentes de los regímenes innegablemente represivos como Corea del Norte o Arabia Saudí. Mercier, invocando una distinción formulada por Sperber, propone que las teorías de la conspiración y las creencias extrañas son *reflexivas* —fruto de la meditación y la teorización consciente—, más que *intuitivas* —las convicciones que sentimos en nuestros huesos—.[41] Se trata de una poderosa distinción, aunque yo la trazo de un modo un tanto

diferente, más próximo al contraste que el psicólogo social Robert Abelson y el comediante George Carlin establecían entre creencias *distales* y *comprobables*.[42]

Los individuos dividen su mundo en dos zonas. Una consta de los objetos físicos que los rodean, las otras personas con las que tratan cara a cara, el recuerdo de sus interacciones, y las reglas y normas que regulan sus vidas. Las personas poseen creencias básicamente precisas acerca de esta zona y razonan racionalmente dentro de ella. En esta zona, creen que existe un mundo real y que las creencias acerca de este son verdaderas o falsas. No tienen elección: esa es la única manera de tener gasolina en el coche, dinero en el banco y a los hijos vestidos y alimentados. Podemos llamarla *mentalidad realista*.

La otra zona es el mundo más allá de la experiencia inmediata: el pasado distante, el futuro incognoscible, los pueblos y lugares lejanos, los corredores remotos del poder, lo microscópico, lo contrafactual, lo metafísico. Las personas pueden abrigar ideas acerca de lo que sucede en estas zonas, pero no tienen manera de comprobarlas y, de todos modos, ello no marca ninguna diferencia discernible en sus vidas. Las creencias en estas zonas son relatos, que pueden ser entretenidos, inspiradores o moralmente edificantes. Si son literalmente «verdaderas» o «falsas» no es la pregunta adecuada. La función de estas creencias es construir una realidad social que cohesione a la tribu o secta y le confiera un propósito moral. Podemos llamarla *mentalidad mitológica*.

Es célebre la sentencia de Bertrand Russell: «Es indeseable creer una proposición cuando no hay fundamento alguno para suponer que sea cierta». La clave para comprender la irracionalidad rampante estriba en reconocer que el enunciado de Russell no es un truismo, sino un manifiesto revolucionario. Durante la mayor parte de la historia y la prehistoria humana, no ha existido ningún fundamento para suponer que las proposiciones acerca de mundos remotos fueran verdaderas. Pero las creencias sobre ellos podían resultar empoderadoras o inspiradoras, y eso las hacía sobradamente deseables.

La máxima de Russell es el lujo de una sociedad tecnológicamente avanzada con ciencia, historia, periodismo y su infraestructura de búsqueda de la verdad, incluidos los registros de archivos, los conjuntos de datos digitales, los instrumentos de alta tecnología y las comunidades de edición, verificación de datos y revisión por pares. Como hijos de la Ilustración, abrazamos el credo radical del realismo universal: sostenemos que *todas* nuestras creencias deberían caer dentro de la mentalidad realista. Nos preocupa que nuestra historia de la creación, nuestras leyendas fundacionales, nuestras teorías de los nutrientes, los gérmenes y las fuerzas invisibles, nuestras concepciones de los poderosos, nuestras sospechas acerca de nuestros enemigos, sean verdaderas o falsas. Ello se debe al hecho de que disponemos de las herramientas para obtener respuestas a estas preguntas, o al menos para asignarles grados justificados de creencia. Y contamos con un Estado tecnocrático que, en teoría, debería poner en práctica estas creencias.

Ahora bien, por muy deseable que sea ese credo, no es la manera natural humana de creer. Al otorgar un mandato imperialista a la mentalidad realista para conquistar el universo de las creencias y empujar la mitología hacia los márgenes, somos *nosotros* los extraños o, como gustan de decir los científicos sociales evolutivos, los WEIRD: occidentales, educados, industrializados, ricos y democráticos.[43*] Al menos lo somos los muy instruidos de entre nosotros, en nuestros mejores momentos. La mente humana está adaptada para comprender las esferas remotas de la existencia mediante una mentalidad mitológica. Esto no se debe al hecho de que descendamos específicamente de los cazadores recolectores del Pleistoceno, sino a que descendemos de personas que no suscribían el ideal ilustrado del realismo universal. El sometimiento de todas nuestras creencias a los juicios de la razón y las evidencias es una destreza antinatural, como la alfabetización y el cálculo, y ha de ser inculcada y cultivada.

Y a pesar de todas las conquistas de la mentalidad realista, la mentalidad mitológica continúa ocupando franjas de territorio en el paisaje de las creencias establecidas. El ejemplo evidente es la religión. Más de 2.000 millones de personas creen que, si alguien no acepta que Jesús es su salvador, será condenado al tormento eterno en el infierno.

Afortunadamente, no dan el siguiente paso lógico y tratan de convertir a la gente al cristianismo a punta de espada por su propio bien, o torturan a los herejes que podrían arrastrar a otros hacia la condenación. Sin embargo, en siglos pasados, cuando las creencias cristianas caían en la zona realista, muchos cruzados, inquisidores, conquistadores y soldados en las guerras de religión hacían exactamente eso. Al igual que el redentor de la pizzería Comet Ping Pong, trataban sus creencias literalmente como certezas. En realidad, aunque muchos profesan creer en un más allá, parecen no tener prisa para dejar este valle de lágrimas por la dicha eterna en el paraíso.

Afortunadamente, la creencia religiosa occidental está aparcada de manera segura en la zona mitológica, donde muchas personas protegen su soberanía. A mediados de la primera década de este siglo, los «nuevos ateos», Sam Harris, Daniel Dennett, Christopher Hitchens y Richard Dawkins, se convirtieron en el blanco de la vituperación no solo de evangelistas celosos de la Biblia, sino también de intelectuales establecidos. Estos feateos (*faithesists*, como los llama el biólogo Jerry Coyne), o creyentes en la creencia (término de Dennett), no rebatían que Dios existiera de hecho.[44] Sugerían que es inapropiado o burdo considerar la existencia de Dios una cuestión de verdad o falsedad. La creencia en Dios es una idea que cae fuera de la esfera de la realidad comprobable.

Otra zona de irrealidad convencional son los mitos nacionales. La mayoría de los países consagran un relato fundacional como parte de su conciencia colectiva. Hubo un tiempo en el que estos eran epopeyas de héroes y dioses, como la *Ilíada*, la *Eneida*, las leyendas artúricas y las óperas wagnerianas. Más recientemente han sido guerras de independencia o luchas anticoloniales. Entre los temas comunes figuran la antigua esencia de la nación definida mediante una lengua, una cultura y una patria; un sueño prolongado y un glorioso despertar; una larga historia de victimización y opresión; y una generación de libertadores y fundadores sobrehumanos. Los guardianes del patrimonio mítico no sienten la necesidad de llegar hasta el fondo de lo realmente acaecido, y pueden estar molestos con los historiadores que lo ubican en la zona realista y sacan a la luz su historia trivial, su identidad construida, sus provocaciones recíprocas con los vecinos y los pies de barro de los padres fundadores.

Aún hay otra zona de creencias ni ciertas del todo ni completamente falsas: la ficción histórica y la historia novelada. Parece pedante señalar que Enrique V no pronunció las conmovedoras palabras el día de San Crispín que Shakespeare le atribuyó. Sin embargo, la obra pretende ser una crónica de acontecimientos reales más que un producto de la imaginación del dramaturgo, y de lo contrario no la disfrutaríamos de la misma manera. Lo mismo sucede con las historias noveladas de guerras y luchas más cercanas, que son, en efecto, noticias falsas situadas en el pasado reciente. Cuando los acontecimientos se acercan demasiado al presente o la ficcionalización reescribe hechos importantes, los historiadores pueden hacer sonar la alarma, como cuando Oliver Stone dio vida a la teoría conspiratoria sobre un asesinato en la película de 1991 *JFK*. En 2020, el columnista Simon Jenkins puso objeciones a la serie de televisión *The Crown [La corona]*, una historia dramatizada sobre la reina Isabel y su familia que se tomaba libertades con muchos de los acontecimientos descritos: «Cuando enciendas el televisor esta noche, imagínate que ves las noticias escenificadas en vez de leídas [...]. Acto seguido la BBC proyecta en la pantalla la declaración de que están “basadas en sucesos verdaderos” y espera que hayamos disfrutado».[45] No obstante, la suya era una voz que clamaba en el desierto. La mayoría de los críticos y espectadores no tenían problema alguno con las falsedades suntuosamente filmadas, y Netflix se negó a publicar una advertencia de que algunas de las escenas eran ficticias (aunque sí publicó una advertencia de contenido sensible sobre bulimia).[46]

La frontera entre la zona realista y la mitológica puede variar con las épocas y la cultura. Desde la Ilustración, las mareas en el Occidente moderno han erosionado la zona mitológica, un cambio histórico que el sociólogo Max Weber denominó el *desencantamiento del mundo*. Pero en las fronteras siempre hay escaramuzas. Las mentiras descaradas y las conspiraciones de la posverdad trumpiana pueden concebirse como un intento de reclamar el discurso político para la tierra de la mitología en lugar de para la tierra de la realidad. Al igual que las tramas de las leyendas, las Escrituras y los dramas, son una especie de teatro; que pueda demostrarse su verdad o su falsedad es algo que no viene al caso.

LA PSICOLOGÍA DE LOS APÓCRIFOS

Una vez que apreciamos que los humanos pueden tener creencias que no consideren fácticamente verdaderas, podemos comenzar a comprender la paradoja de la racionalidad: cómo es posible que un animal racional abrace tantos disparates. No se trata de que los teóricos de la conspiración, las personas que comparten noticias falsas y los consumidores de pseudociencia construyan *siempre* sus mitos como mitológicos. A veces, sus creencias cruzan la línea de la realidad con resultados trágicos, como en el Pizzagate, los antivacunas y el culto de la Puerta del Cielo, cuyos treinta y nueve devotos se suicidaron en 1997 como preparación para que se llevase sus almas una nave espacial siguiendo el cometa Hale-Bopp. Pero las predisposiciones en la naturaleza humana pueden combinarse con la veracidad mitológica para hacer fáciles de digerir las creencias extrañas. Examinemos tres géneros.

La pseudociencia, los fenómenos paranormales y el curanderismo movilizan nuestras intuiciones cognitivas más profundas.[47] Somos intuitivamente dualistas y sentimos que las mentes pueden existir aparte de los cuerpos.[48] Esto es algo natural en nosotros, y no solo porque no veamos las redes neuronales que subyacen a las creencias y los deseos de nosotros mismos y de los demás. Muchas de nuestras experiencias sugieren realmente que la mente no está amarrada al cuerpo, incluidos los sueños, los trances, las experiencias extracorpóreas y la muerte. No supone un gran salto concluir que las mentes pueden estar en comunión con la realidad y las unas con las otras sin necesidad de un soporte físico. Y así surgen la telepatía, la clarividencia, las almas, los fantasmas, la reencarnación y los mensajes desde el más allá.

Somos, asimismo, intuitivamente esencialistas, y sentimos que los seres vivos contienen sustancias invisibles que les confieren su forma y sus poderes.[49] Estas intuiciones inspiran a la gente a investigar los seres vivos en busca de semillas, drogas y venenos. Pero esta mentalidad también hace creer a la gente en la homeopatía, los remedios herbales, las purgas y sangrías, y el rechazo de adulterantes extraños como las vacunas y los alimentos genéticamente modificados.

Y somos intuitivamente teleólogos.[50] Al igual que nuestros planes y artefactos están diseñados con un propósito, propendemos a pensar que también lo está la complejidad del mundo viviente y no viviente. Así pues, somos receptivos al creacionismo, la astrología, la sincronicidad y la creencia mística en que todo sucede por una razón.

Supuestamente, una educación científica ha de sofocar estas intuiciones primitivas, pero su alcance es limitado por varios motivos. Uno de ellos es que no resulta fácil renunciar a las creencias que son sagradas para una facción religiosa o cultural, como el creacionismo, el alma y un propósito divino, y estas pueden guardarse dentro de la zona mitológica de los individuos. Otro es que, incluso entre la gente muy culta, la comprensión científica es superficial. Pocas personas pueden explicar por qué el cielo es azul o por qué cambian las estaciones, por no hablar de la genética de poblaciones o la inmunología viral. En lugar de ello, los individuos instruidos confían en la comunidad científica y universitaria: su consenso es suficiente para ellos.[51]

Desgraciadamente, para mucha gente la frontera entre la comunidad científica y la pseudocientífica es borrosa. Para la mayoría de las personas, lo más cercano a la ciencia en su vida es su médico, y muchos médicos son más curanderos populares que expertos en ensayos clínicos aleatorizados. De hecho, algunos de los doctores famosos que aparecen en los programas matinales de entrevistas son charlatanes que promocionan de forma exuberante engañifas de la nueva era. Los documentales e informativos de la televisión convencional también pueden difuminar las líneas y dramatizar crédulamente teorías especulativas como los antiguos astronautas y los videntes que luchan contra la delincuencia.[52]

En realidad, los divulgadores científicos serios deben cargar con parte de la culpa por no equipar a la gente con la comprensión profunda que tornaría increíble la pseudociencia a simple vista. A menudo, la ciencia se presenta en las escuelas y en los museos como otra forma de magia oculta, con criaturas exóticas, sustancias químicas pintorescas e ilusiones llamativas. Los principios fundacionales, como que el universo no tiene objetivos relacionados con las preocupaciones humanas, que todas las interacciones físicas están gobernadas por unas cuantas fuerzas

fundamentales, que los cuerpos vivientes son intrincadas máquinas moleculares y que la mente es la actividad de procesamiento de la información del cerebro, nunca se expresan con claridad, quizá porque se ofenderían las sensibilidades religiosas y morales. No debería sorprendernos que lo que la gente saque en limpio de la educación científica sea un batiburrillo sincrético, donde la gravedad y el electromagnetismo coexisten con el *psi*, el *qi*, el karma y la sanación con cristales.

PARA ENTENDER PATRAÑAS VIRALES tales como las leyendas urbanas, los titulares de los tabloides y las noticias falsas, hemos de recordar que son extraordinariamente entretenidas. Tratan temas de sexo, violencia, venganza, peligro, fama, magia y tabú, que siempre han despertado el interés de los mecenas de las artes, tanto altas como bajas. Un titular falso como «Agente del FBI sospechoso de las filtraciones de correos electrónicos de Hillary hallado muerto en un aparente asesinato-suicidio» sería una excelente trama de un *thriller*. Un reciente análisis cuantitativo de los contenidos de las noticias falsas concluía que «las mismas características que tornan culturalmente atractiva cualquier leyenda urbana, ficción y, de hecho, cualquier relato, operan también para cualquier información falsa *online*».[53]

El entretenimiento se derrama con frecuencia en géneros de comedia, incluidas la bufonada, la sátira y la farsa: «Empleado de morgue incinerado por error mientras se echaba una siesta»; «Donald Trump acaba con los tiroteos escolares prohibiendo las escuelas»; «Bigfoot retiene a un leñador como esclavo sexual». QAnon cae en otro género de entretenimiento, el juego multiplataforma de realidad alternativa.[54] Los seguidores analizan pistas crípticas dejadas periódicamente por Q (el hipotético soplón gubernamental), examinan colaborativamente sus hipótesis y ganan fama en internet compartiendo sus descubrimientos.

No es de extrañar que la gente busque toda suerte de entretenimientos. Lo que nos sorprende es que estas obras de arte se conviertan en una afirmación fáctica. No obstante, nuestra intranquilidad por la confusión entre hecho y ficción no es una reacción humana universal, particularmente cuando afecta a zonas remotas de la experiencia inmediata, como lugares

lejanos o las vidas de los ricos y poderosos. Del mismo modo que los mitos religiosos y nacionales se arraigan en la corriente dominante cuando se siente que proporcionan elevación moral, las noticias falsas pueden volverse virales cuando sus propagadores piensan que está en juego un valor superior, como reforzar la solidaridad en el seno de su bando y recordar a los camaradas la perfidia del bando contrario. A veces la moraleja no es siquiera una estrategia política coherente, sino un sentido de superioridad moral: la impresión de que las clases sociales rivales y las poderosas instituciones respecto de las cuales se sienten alienados los propagadores son decadentes y corruptas.

LAS TEORÍAS DE LA CONSPIRACIÓN, por su parte, florecen porque los humanos siempre han sido vulnerables a las conspiraciones reales.[55] Para los pueblos cazadores recolectores, todas las precauciones son pocas. La forma más letal de guerra entre los pueblos tribales no es la batalla campal, sino la emboscada sigilosa y el ataque antes del amanecer.[56] El antropólogo Napoleon Chagnon escribe que los yanomamis amazónicos tienen la palabra *nomohori*, «truco ruin», para los actos de traición como invitar a los vecinos a un banquete y luego masacrarlos en el momento justo. Los complots de las coaliciones enemigas son diferentes de otros peligros, como los depredadores y los rayos, porque estas despliegan su ingenio para atravesar las defensas de sus objetivos y cubrir sus propios rastros. La única salvaguarda contra este subterfugio de intriga y misterio consiste en superar al otro en astucia con carácter preventivo, lo cual puede conducir a intrincadas series de conjeturas y a la negativa a creer a pies juntillas los hechos evidentes. En términos de detección de señales, el coste de pasar por alto una conspiración real es más elevado que el de una falsa alarma ante una sospecha. Esto requiere fijar el sesgo hacia el extremo de la escala de disparar con facilidad más que hacia el de disparar con comedimiento, adaptándonos para intentar enterarnos de las posibles conspiraciones incluso con evidencias endebles.[57]

Aun en nuestros días siguen existiendo las conspiraciones grandes y pequeñas. Un grupo de empleados puede reunirse a espaldas de un colega impopular para recomendar su despido, un Gobierno o un grupo insurgente

puede planear clandestinamente un golpe de Estado, una invasión o un sabotaje. Las teorías de la conspiración, al igual que las leyendas urbanas y las noticias falsas, logran convertirse en rumores, y los rumores son la materia prima de la conversación. Los estudios sobre los rumores muestran que estos tienden a transmitir amenazas y peligros, y que confieren un aura de sabiduría a quien los difunde. Y tal vez sorprendentemente, cuando circulan entre personas con un interés creado en su contenido, como en el caso de los lugares de trabajo, suelen ser correctos.[58]

Así pues, en la vida cotidiana hay incentivos para ser un centinela que advierte de las amenazas ocultas, o un repetidor que propaga sus avisos. El problema estriba en que los medios sociales y los medios de comunicación permiten que los rumores se propaguen por redes de personas que no tienen ningún interés en su verdad. Estas consumen los rumores como entretenimiento y afirmación, más que como autoprotección, y carecen del interés y de los medios para hacer un seguimiento de ellos. Por las mismas razones, los creadores y propagadores no sufren ningún daño en su reputación por equivocarse. Sin estas comprobaciones de veracidad, los rumores en los medios sociales, a diferencia de los rumores en los lugares de trabajo, son con más frecuencia *incorrectos* que correctos. Mercier sugiere que la mejor forma de impedir la propagación de noticias dudosas es presionar a los propagadores para que actúen en consecuencia: llamar a la policía en lugar de dejar una reseña de una estrella.

La clave restante para comprender el atractivo de las creencias extrañas es examinar las creencias mismas con microscopio. La evolución no solo funciona en los cuerpos y los cerebros, sino también en las ideas. Un *meme*, tal como lo definió Richard Dawkins cuando acuñó el término, no es una fotografía con un pie que circula por internet, sino una idea que ha sido moldeada por generaciones que la han ido compartiendo hasta tornarse muy compartible.[59] Entre los ejemplos figuran las melodías pegadizas que la gente no puede dejar de tararear o las historias que se siente obligada a transmitir. Al igual que los organismos desarrollan adaptaciones que los protegen de ser devorados, las ideas pueden desarrollar adaptaciones que las protegen de ser refutadas. El ecosistema intelectual está repleto de estas ideas invasivas:[60] «Los designios de Dios

son inescrutables», «La negación es un mecanismo de defensa del ego», «Los poderes psíquicos son inhibidos por la exploración escéptica», «Si no denuncias a esta persona por racista, eso demuestra que tú eres racista», «Todo el mundo es egoísta, porque ayudar a los demás hace que te sientas bien», y, por supuesto, «La falta de pruebas de esta conspiración demuestra lo diabólica que es». Las teorías conspiratorias, por su propia naturaleza, están adaptadas para ser propagadas.

Reafirmar la racionalidad

Comprender no implica perdonar. Podemos ver por qué los humanos dirigen sus razonamientos hacia conclusiones que redundan en beneficio de sí mismos o de sus sectas, y por qué distinguen una realidad en la que las ideas son verdaderas o falsas de una mitología en la que las ideas son entretenidas o inspiradoras, sin conceder que estas cosas sean buenas. No son buenas. La realidad es aquello que, cuando le aplicas razonamientos motivados, o *de mi lado*, o mitológicos, no desaparece. Las creencias falsas sobre las vacunas, las medidas de salud pública y el cambio climático amenazan el bienestar de miles de millones de personas. Las teorías conspirativas incitan al terrorismo, los pogromos, las guerras y el genocidio. La corrosión de los estándares de la verdad socava la democracia y despeja el terreno para la tiranía.

No obstante, pese a todas las vulnerabilidades de la razón humana, nuestra imagen del futuro no tiene por qué ser un bot tuiteando eternamente noticias falsas. El arco del conocimiento es largo y se dobla hacia la racionalidad. No deberíamos perder de vista cuánto prolifera la racionalidad ahí afuera. Hoy en día son pocas las personas de los países desarrollados que creen en los hombres lobo, los sacrificios animales, las sangrías, los miasmas, el derecho divino de los líderes o los augurios en los eclipses y los cometas, aunque todas estas creencias estaban a la orden del día en siglos pasados. Ninguna de las treinta mil falsedades de Trump implicaban fuerzas ocultas o paranormales, y cada una de estas fuerzas es rechazada por una mayoría de los estadounidenses.[61] Aunque unos cuantos asuntos científicos caldean los ánimos religiosos o políticos, la mayoría no: hay facciones que

desconfían de las vacunas, pero no de los antibióticos; del cambio climático, pero no de la erosión costera.[62] A pesar de sus sesgos partidistas, la mayoría de las personas juzgan con muy buen criterio la veracidad de los titulares y, cuando se les presentan correcciones claras y fiables de una afirmación falsa, cambian de opinión, independientemente de sus simpatías o antipatías políticas.[63]

Tenemos asimismo una cabeza de puente de la racionalidad en el estilo cognitivo llamado *apertura mental activa*, especialmente en el subtipo denominado *apertura a las evidencias*.[64] Este es el credo de Russell en virtud del cual las creencias deberían estar basadas en buenas razones. Es un rechazo del razonamiento motivado; un compromiso para ubicar todas las creencias dentro de la zona realista; un respaldo a la declaración atribuida a John Maynard Keynes: «Cuando los hechos cambian, cambio de opinión. ¿Y qué hace usted, señor?».[65] El psicólogo Gordon Pennycook y sus colegas midieron la actitud haciendo que los sujetos rellenasen un cuestionario con ítems como estos, en los que la respuesta entre paréntesis aumenta la puntuación de la apertura:[66]

> La gente debería tener siempre en cuenta las evidencias que van en contra de sus creencias (DE ACUERDO).
>
> Ciertas creencias son demasiado importantes como para abandonarlas por muy sólidos que sean los argumentos en contra de ellas (EN DESACUERDO).
>
> Las creencias deberían revisarse siempre en respuesta a las nuevas informaciones o evidencias (DE ACUERDO).
>
> Nadie puede disuadirme de algo que sé que es cierto (EN DESACUERDO).
>
> Creo que la fidelidad a mis ideales y principios es más importante que la «apertura mental» (EN DESACUERDO).

En una muestra de usuarios de internet estadounidenses, en torno a una quinta parte de los encuestados dicen ser insensibles a las evidencias, pero la mayoría aspira al menos a estar abierta a ellas. Las personas abiertas a las evidencias se resisten a las creencias extrañas. Rechazan las teorías conspirativas, la brujería, la astrología, la telepatía, los augurios y al monstruo del lago Ness, junto con un Dios personal, el creacionismo, la Tierra joven, el vínculo entre las vacunas y el autismo, y la negación del cambio climático antropogénico.[67] Confían más en el Gobierno y en la ciencia. Y tienden a defender posiciones políticas más liberales en temas

como el aborto, el matrimonio homosexual, la pena de muerte y la aversión a la guerra, generalmente en las mismas direcciones en las que ha estado avanzando el mundo en su conjunto[68] (los autores advierten, sin embargo, que las correlaciones con el conservadurismo son complicadas).

La apertura a las evidencias se correlaciona con la reflexión cognitiva (la capacidad de pensar dos veces y no dejarse engañar por las preguntas capciosas, que conocimos en el capítulo 1) y con la resistencia a muchas de las ilusiones cognitivas, sesgos y falacias que vimos en los capítulos 3 a 9.[69] Este conjunto de buenos hábitos cognitivos, que Stanovich denomina *cociente de racionalidad* (un juego con el cociente de inteligencia, CI), se correlaciona con la inteligencia en bruto, si bien de manera imperfecta: las personas inteligentes pueden ser cerradas de mente e impulsivas, y las más torpes pueden ser abiertas y reflexivas. Además de resistirse a las creencias extrañas, los individuos reflexivos son mejores detectando noticias falsas y rechazando sandeces pseudoprofundas como «el significado oculto transforma la belleza abstracta sin igual».[70]

Si pudiéramos añadir algo al agua potable que volviera a todo el mundo más abierto y reflexivo, se esfumaría la crisis de irracionalidad. A falta de ello, consideremos un amplio repertorio de políticas y normas que podrían fortalecer los sistemas inmunes cognitivos en nosotros mismos y en nuestra cultura.[71]

Lo más radical sería una valorización de la propia norma de la racionalidad. Ahora bien, no podemos imponer valores desde arriba en mayor medida en que podemos dictar cualquier cambio cultural que dependa de millones de elecciones individuales, como los tatuajes o el argot. Pero las normas pueden cambiar con el tiempo, como la disminución de las calumnias étnicas, la basura tirada en la calle y los chistes sobre esposas, cuando los reflejos de aprobación y desaprobación tácitas proliferan por las redes sociales. Así pues, cada uno de nosotros podemos poner nuestro granito de arena sonriendo o frunciendo el ceño ante los hábitos racionales o irracionales. Sería estupendo ver a los individuos ganar puntos por reconocer la incertidumbre en sus creencias, cuestionar los dogmas de su secta política y cambiar de parecer cuando cambian los hechos, en lugar de ser adalides inquebrantables de los dogmas de su

camarilla. Inversamente, podría suponer un paso en falso mortificante otorgar una importancia excesiva a las anécdotas, confundir la correlación con la causalidad o cometer una falacia informal como la culpa por asociación o el argumento de autoridad. La Rationality Community o Comunidad de la Racionalidad se identifica a sí misma mediante estas normas, pero estas deberían ser costumbre para toda la sociedad, más que el pasatiempo de un club de entusiastas.[72]

Aunque resulta difícil pilotar el portaaviones que constituye una sociedad entera, las instituciones particulares pueden tener puntos de presión sobre los que podrían actuar los líderes y activistas expertos. Los órganos legislativos están poblados en buena medida por abogados, cuya meta profesional es la victoria más que la verdad. Recientemente algunos científicos han comenzado a infiltrarse en las cámaras, y podrían tratar de difundir entre sus colegas el valor de la resolución de problemas basada en las evidencias. Los defensores de cualquier política harían bien en no etiquetarla con simbolismo sectario; algunos expertos climáticos, por ejemplo, lamentaron que Al Gore se convirtiera en la cara visible del activismo contra el cambio climático a comienzos de la década de 2000, pues eso lo encasillaba como una causa de la izquierda, dando a la derecha una excusa para oponerse a él.

Entre los políticos, los dos principales partidos de Estados Unidos adolecen de un marcado *sesgo de mi lado*, pero la culpa no es simétrica. Incluso antes de la llegada al poder de Trump, algunos reflexivos baluartes republicanos denigraban su propia organización como «el partido de los estúpidos» por su antintelectualismo y su hostilidad hacia la ciencia.[73] Desde entonces, muchos otros se han sentido horrorizados por su aquiescencia ante las mentiras y los troleos maniacos de Trump, cuya estrategia, en las admirativas palabras del antiguo estratega Steve Bannon, consistía en «inundar la zona con mierda».[74] Con la derrota de Trump, las cabezas racionales de la derecha deberían tratar de restaurar en la política estadounidense un sistema con dos partidos que difieren acerca de sus políticas respectivas, más que en lo que atañe a la existencia de los hechos y a la verdad.

No estamos indefensos ante la avalancha de desinformación de la *posverdad*. Aunque las mentiras son tan viejas como el lenguaje, también lo son las defensas contra ellas; como señala Mercier, sin esas defensas el lenguaje nunca podría haber evolucionado.[75] También las sociedades se protegen contra las inundaciones de mierda: a los mentirosos descarados se los considera responsables, se les imponen sanciones legales y su reputación queda dañada. Estas salvaguardas se están desplegando tardíamente. En una sola semana a principios de 2021, las compañías fabricantes de las máquinas y el *software* de votación mencionadas en las teorías conspirativas de Trump demandaron por difamación a algunos miembros de su equipo jurídico; Trump fue bloqueado en Twitter por violar su política contra la incitación a la violencia; un senador mendaz que promovió en el Congreso la teoría conspirativa del robo electoral perdió un importante contrato editorial; y el editor de la revista *Forbes* anunció: «Que el mundo empresarial lo sepa: contraten a cualquiera de los fabulistas de Trump y *Forbes* asumirá que todo aquello de lo que hable su empresa o su firma es mentira».[76]

Dado que nadie puede saberlo todo, y la mayoría de las personas no saben casi nada, la racionalidad consiste en externalizar el conocimiento a instituciones especializadas en crearlo y compartirlo, principalmente el mundo académico, las unidades de investigación públicas y privadas, y la prensa.[77] Esa confianza es un precioso recurso que no debería ser derrochado. Aunque la confianza en la ciencia haya permanecido estable durante décadas, la confianza en las universidades se está hundiendo.[78] Un motivo importante para la desconfianza es el asfixiante monocultivo izquierdista de las universidades, que castigan a los estudiantes y profesores que cuestionan los dogmas sobre el género, la raza, la cultura, la genética, el colonialismo, y la identidad y la orientación sexual. Las universidades se han convertido en el hazmerreír por sus ataques contra el sentido común (como cuando un profesor fue suspendido recientemente por mencionar la muletilla china *ne ga*, porque a algunos estudiantes les recordaba el insulto racista *nigger* («negrata»).[79] En varias ocasiones he recibido cartas de personas que me preguntaban por qué habrían de confiar en el consenso científico sobre el cambio climático, toda vez que este procede de

instituciones que no toleran el disenso. Esta es la razón por la que las universidades tienen la responsabilidad de garantizar la credibilidad de la ciencia y la erudición, comprometiéndose con la diversidad de puntos de vista, la libre investigación, el pensamiento crítico y la apertura mental activa.[80]

La prensa, perennemente empatada con el Congreso como la institución estadounidense que inspira menos confianza, tiene que desempeñar asimismo un papel especial en la infraestructura de la racionalidad.[81] Al igual que las universidades, los sitios de noticias y de opinión deberían ser modelos de diversidad de puntos de vista y pensamiento crítico. Y, como he argüido en el capítulo 4, deberían fortalecer también su competencia aritmética y su capacidad de interpretación de los datos, y ser conscientes de las ilusiones estadísticas instiladas por la caza sensacionalista de la anécdota. Hay que reconocer que los periodistas se han vuelto más conscientes de cómo pueden ser engañados por políticos insinceros y contribuir a los miasmas de la posverdad, y han comenzado a implementar contramedidas como verificar los hechos, etiquetar las afirmaciones falsas y no repetirlas, exponer los hechos afirmativamente en lugar de en forma negativa, corregir errores con claridad y rapidez, y evitar un falso equilibrio entre expertos y excéntricos.[82]

Las instituciones educativas, desde las escuelas de enseñanza primaria hasta las universidades, podrían hacer que el pensamiento estadístico y crítico ocupara una parte mayor de sus currículos. Al igual que la alfabetización y el cálculo ocupan un lugar preferente en la escolarización porque son un prerrequisito para todo lo demás, las herramientas de la lógica, la probabilidad y la inferencia causal recorren toda clase de conocimiento humano. La racionalidad debería convertir en cuarteto el trío clásico de la lectura, la escritura y la aritmética. Sin lugar a dudas, la mera formación en probabilidad no proporciona inmunidad de por vida ante las falacias estadísticas. Los estudiantes la olvidan en cuanto se examinan y venden sus libros de texto, e incluso cuando recuerdan la materia, casi ninguno da el salto de los principios abstractos a los escollos cotidianos.[83] Pero que los cursos y los videojuegos bien diseñados (que señalen los

sesgos cognitivos: la falacia del jugador, los costos hundidos o irrecuperables, el sesgo de confirmación y demás) reten a los estudiantes a detectarlos en escenarios realistas, reformulen los problemas en formatos amigables para la mente y les doten de retroalimentación inmediata sobre sus errores puede entrenarlos de verdad para evitar las falacias fuera del aula.[84]

LA RACIONALIDAD ES UN BIEN PÚBLICO, y un bien público prepara el terreno para una tragedia de los comunes. En la tragedia de los comunes de la racionalidad, el razonamiento motivado en beneficio de uno mismo y de su bando crea una oportunidad de aprovecharse de nuestra comprensión colectiva.[85] Cada uno de nosotros tenemos un motivo para preferir nuestra verdad, pero todos juntos salimos ganando con *la* verdad.

Las tragedias de los comunes pueden mitigarse con normas informales en las que los miembros de una comunidad vigilen las tierras de pastoreo o los caladeros reconociendo a los buenos ciudadanos y estigmatizando a los explotadores.[86] Las sugerencias que he hecho hasta ahora pueden, a lo sumo, fortalecer a los razonadores individuales e inculcar la norma de que el razonamiento sensato es una virtud. Pero los bienes comunes han de protegerse asimismo con incentivos: recompensas que hagan que a los razonadores les interese respaldar las ideas con la mayor justificación. Obviamente no podemos implementar un impuesto a las falacias, pero los integrantes de la colectividad pueden acordar unas reglas que reorienten los incentivos hacia la verdad.

He mencionado que las instituciones exitosas de la racionalidad jamás dependen de la brillantez de un individuo, pues ni siquiera el más racional de nosotros está libre de sesgos. En lugar de ello, cuentan con canales de retroalimentación y agregación de conocimientos que hacen el todo más inteligente que cualquiera de sus partes.[87] Entre ellos figuran la revisión por pares en el mundo académico, la comprobabilidad en la ciencia, la verificación de hechos y la edición en el periodismo, los controles y contrapesos en la gobernanza y los procedimientos contradictorios en el sistema judicial.

Los nuevos medios de comunicación de cada época abren un salvaje Oeste de apócrifos y robos de la propiedad intelectual hasta que se ponen en marcha las contramedidas al servicio de la verdad.[88] Eso es lo que sucedió en el pasado con los libros y luego con los periódicos, y lo que está ocurriendo hoy en día con los medios digitales. Los medios de comunicación pueden llegar a ser o bien crisoles de conocimientos o bien cloacas de bobadas, dependiendo de su estructura de incentivos. El sueño en los albores de la era de internet —según el cual, al proporcionar a todo el mundo una plataforma, surgiría una nueva Ilustración— parece hoy digno de vergüenza, ahora que vivimos con los bots, los troles, las discusiones acaloradas e improductivas, las noticias falsas, las vergonzosas turbas de Twitter y el ciberacoso. En la medida en que la aceptación en la plataforma digital consista en los «me gusta», lo compartido, los clics y los globos oculares, no tenemos ningún motivo para pensar que esto vaya a fomentar la racionalidad o la verdad. Wikipedia, en cambio, aunque no es infalible, se ha convertido en un recurso asombrosamente preciso, pese a ser libre y descentralizado. Esto se debe a que implementa de forma intensiva la corrección de errores y el control de calidad, respaldados por los «pilares» destinados a marginar los sesgos de mi lado.[89] Entre estos figuran la verificabilidad, un punto de vista neutral, el respeto y el civismo, y la misión de ofrecer conocimientos objetivos. Como el sitio proclama: «Wikipedia no es una tribuna improvisada, una plataforma publicitaria, una editorial de autopublicación [ni] un experimento de anarquía o democracia».[90]

Mientras escribo estas páginas, esos experimentos colosales de anarquía y democracia, las plataformas de los medios sociales, han comenzado a cobrar conciencia de la tragedia de los comunes de la racionalidad, despertados por dos alarmas activadas en 2020: la información errónea y la desinformación acerca de la pandemia de la COVID, y las amenazas a la integridad de las elecciones presidenciales estadounidenses. Las plataformas han afinado sus algoritmos para dejar de recompensar las falsedades peligrosas, han insertado etiquetas de advertencia y enlaces de verificación de hechos, y han sofocado las dinámicas desbocadas que pueden viralizar los contenidos tóxicos e

introducir a la gente en madrigueras de conejo extremistas. Es demasiado pronto para saber qué funcionará y qué no.[91] Es evidente que estos esfuerzos deberían redoblarse, con vistas a reformar la perversa estructura de incentivos que recompensa la notoriedad sin premiar en modo alguno la verdad.

Ahora bien, así como es probable que se culpe en exceso a los medios sociales por la irracionalidad partidista, sus retoques algorítmicos no bastarán para repararla. Deberíamos ser creativos a la hora de cambiar las reglas en otros ámbitos, de suerte que se otorgue una ventaja a la verdad desinteresada sobre el sesgo de mi lado. En el periodismo de opinión, cabría juzgar a los expertos en función de la exactitud de sus pronósticos, en lugar de su capacidad de sembrar temor y odio o enardecer los ánimos de una facción.[92] En política, medicina, mantenimiento del orden público y otras especialidades, la evaluación basada en evidencias debería ser una práctica generalizada, no específica.[93] Y en el Gobierno, las elecciones, que pueden sacar lo peor del razonamiento, podrían complementarse con la democracia deliberativa, como paneles de ciudadanos encargados de recomendar una política determinada.[94] Este mecanismo pone en práctica el descubrimiento de que, en grupos de razonadores cooperativos pero intelectualmente diversos, suele triunfar la verdad.[95]

El razonamiento humano tiene sus falacias, sus sesgos y su complacencia en la mitología. Pero la explicación última de la paradoja de cómo puede nuestra especie ser tan racional y tan irracional no es un error en nuestra programación cognitiva. Reside en la dualidad del yo y el otro: nuestra capacidad de razonamiento está orientada por nuestros motivos y limitada por nuestros puntos de vista. Vimos en el capítulo 2 que el núcleo de la moralidad es la imparcialidad: la reconciliación de nuestros propios intereses egoístas con los de los demás. Así pues, la imparcialidad es también el núcleo de la racionalidad: una reconciliación de nuestras ideas sesgadas e incompletas con una comprensión de la realidad que trascienda a cualquiera de nosotros. Por tanto, la racionalidad no es solo una virtud cognitiva, sino también moral.

CAPÍTULO 11

Por qué es importante la racionalidad

Comenzar a razonar es como poner el pie en una escalera mecánica que asciende hasta perderse de vista. Una vez que damos el primer paso, la distancia que recorreremos es independiente de nuestra voluntad y no podemos saber de antemano dónde acabaremos.

PETER SINGER[1]

Ofrecer razones por las que es importante la racionalidad se asemeja un poco a soplar en tus velas o a levantarte tirando de las trabillas de tus botas: no puede funcionar a menos que aceptes primero la regla básica de que la racionalidad es la manera de decidir lo que importa. Afortunadamente, como hemos visto en el capítulo 2, todos aceptamos la primacía de la razón, al menos tácitamente, tan pronto como discutimos este asunto, o cualquier otro, en lugar de obligar a asentir por la fuerza. Ha llegado el momento de subir las apuestas y preguntar si la aplicación consciente de la razón mejora realmente nuestras vidas y hace del mundo un lugar mejor. Debería hacerlo, dado que la realidad está gobernada por las leyes lógicas y físicas en lugar de las artes diabólicas y la magia. Pero ¿de veras sufren daños los individuos a causa de sus falacias, y les irían mejor sus vidas si las identificaran y pensaran la forma de escapar de ellas? ¿O son los instintos y las corazonadas mejores guías para las decisiones vitales que la reflexión, con sus riesgos de pensar en exceso y de racionalización?

Cabe hacer las mismas preguntas acerca del bienestar del mundo. ¿Es el progreso una historia de resolución de problemas, impulsada por los filósofos que diagnostican los males, y por los científicos y responsables políticos que encuentran los remedios? ¿O es el progreso una historia de lucha, en la que los oprimidos se levantan y vencen a sus opresores?[2] En capítulos anteriores hemos aprendido a desconfiar de las falsas dicotomías y las explicaciones monocausales, por lo que las respuestas a estas preguntas no serán solo lo uno o lo otro. No obstante, explicaré por qué creo que ejercitar nuestra razón divina en lugar de dejarla «enmohecerse por falta de uso» puede conducirnos a una vida mejor y a un mundo mejor.

La racionalidad en nuestras vidas

¿Son las falacias e ilusiones exhibidas en los capítulos precedentes meras respuestas incorrectas a difíciles problemas matemáticos? ¿Son rompecabezas, trampas, preguntas capciosas, curiosidades de laboratorio?

¿O puede el razonamiento defectuoso causar auténtico daño, con la implicación de que el pensamiento crítico podría protegernos de nuestros peores instintos cognitivos?

Ciertamente, muchos de los sesgos que hemos explorado parecerían sancionados por la realidad, con toda su indiferencia hacia nuestras creencias irracionales.[3] Descontamos el futuro con cortedad de miras, pero este siempre llega, menos las grandes recompensas que sacrificamos a cambio de la rápida satisfacción. Tratamos de recuperar los costos hundidos, de modo que prolongamos demasiado las malas inversiones, las malas películas y las malas relaciones. Evaluamos el peligro en función de la disponibilidad, y evitamos así los seguros aviones, que cambiamos por los peligrosos coches, que conducimos mientras enviamos mensajes de texto. Malinterpretamos la regresión a la media, de manera que buscamos explicaciones ilusorias para los éxitos y los fracasos.

Al ocuparnos del dinero, nuestro punto ciego para el crecimiento exponencial nos lleva a ahorrar demasiado poco para la jubilación y a endeudarnos demasiado con nuestras tarjetas de crédito. Nuestra incapacidad de descontar con precisión *post hoc* y nuestra confianza fuera de lugar en los expertos en las fórmulas actuariales nos conducen a invertir en fondos costosamente gestionados que obtienen peores resultados que los fondos indexados simples. Nuestra dificultad con la utilidad esperada nos tienta con seguros y juegos que nos dejan en peor situación a largo plazo.

En lo tocante a nuestra salud, nuestra dificultad con el razonamiento bayesiano puede aterrorizarnos al sobreinterpretar un resultado positivo en la prueba de una enfermedad común. Podemos ser convencidos o disuadidos de la cirugía dependiendo de la elección de palabras en las que se formulan los riesgos en lugar del equilibrio de riesgos y beneficios. Nuestras intuiciones relativas a las esencias nos llevan a rechazar las vacunas que salvan vidas y a abrazar un curanderismo peligroso. Las correlaciones ilusorias, así como la confusión de la correlación con la causalidad, nos llevan a aceptar diagnósticos y tratamientos carentes de valor de médicos y psicoterapeutas. La incapacidad de sopesar riesgos y recompensas nos hace correr riesgos estúpidos con nuestra seguridad y nuestra felicidad.

En el ámbito legal, la ceguera a la probabilidad puede llevar a jueces y jurados a cometer errores judiciales mediante vívidas conjeturas y probabilidades *post hoc*. La incapacidad de apreciar el equilibrio entre aciertos y falsas alarmas los conduce a castigar a muchos inocentes con el fin de condenar a pocos más de los culpables.

En muchos de estos casos, los profesionales son tan vulnerables a la locura como sus pacientes y clientes, lo cual demuestra que la inteligencia y la pericia no proporcionan inmunidad contra las infecciones cognitivas. Las ilusiones clásicas se han encontrado en trabajadores sanitarios, abogados, inversores, brókeres, escritores deportivos, economistas y meteorólogos, todos los cuales manejan cifras en sus especialidades respectivas.[4]

Estas son algunas de las razones para creer que los fracasos de la racionalidad tienen consecuencias en el mundo. ¿Puede cuantificarse el daño? El activista del pensamiento crítico Tim Farley intentó hacerlo en su sitio web y en su canal de Twitter, que lleva el nombre de la frecuente pregunta What’s the Harm? (¿Cuál es el daño?).[5] Huelga decir que Farley no tenía manera de responderla con precisión, pero trataba de abrir los ojos de la gente a la enormidad del daño causado por los errores del pensamiento crítico enumerando todos los casos autentificados que fue capaz de encontrar. Desde 1970 hasta 2009, pero principalmente en la última década de ese rango, documentó 368.379 personas muertas, más de 300.000 heridas y 2.800 millones de dólares en daños económicos por meteduras de pata en el pensamiento crítico. Entre estos figuraban personas que se mataban o mataban a sus hijos al rechazar tratamientos médicos convencionales o al usar curas herbales, homeopáticas, holísticas y otros remedios de curanderos; suicidios masivos de miembros de cultos apocalípticos; asesinatos de brujas, hechiceros y personas «malditas»; víctimas inocentes a las que les birlan sus ahorros los videntes, astrólogos y demás charlatanes; malhechores y justicieros arrestados por actuar movidos por delirios conspirativos; y pánicos económicos resultantes de supersticiones y falsos rumores. He aquí unos cuantos tuits de 2018-2019:

> ¿Cuál es el daño de las teorías de la conspiración? El FBI identifica a los «extremistas domésticos impulsados por conspiraciones» como una nueva amenaza de terror interno.

> ¿Cuál es el daño de recibir asesoramiento en materia de salud de un #herborista? Un niño de trece años murió después de que le dijeran que no tomara insulina. El herborista va ahora camino de la cárcel.

> ¿Cuál es el daño en una #iglesia de sanación por la fe? Ginnifer luchó por su vida durante cuatro horas. Travis Mitchell, su padre, «le impuso las manos» y la familia se turnaba rezando mientras ella luchaba por respirar y cambiaba de color. «Supe que estaba muerta cuando dejó de gritar», dijo Mitchell.
>
> ¿Cuál es el daño de creer en seres sobrenaturales? Los aldeanos de Sumatra mataron un tigre en peligro de extinción porque pensaron que era un *siluman* que cambia de formas.
>
> ¿Cuál es el daño de acudir a un #vidente? «Vidente» de Maryland condenado por estafar a sus clientes trescientos cuarenta mil dólares.

Como Farley sería el primero en apreciar, ni siquiera miles de anécdotas pueden demostrar que sucumbir a los sesgos irracionales provoca más daño que superarlos. Necesitamos cuando menos un grupo de comparación, a saber, los efectos de instituciones regidas por la razón tales como la medicina, la ciencia y el Gobierno democrático. Este es el tema de la siguiente sección.

Sí que contamos con un estudio de los efectos de la toma de decisiones racionales sobre las consecuencias en la vida. Los psicólogos Wändi Bruine de Bruin, Andrew Parker y Baruch Fischhoff desarrollaron una medición de la competencia en razonamiento y toma de decisiones (como el cociente de racionalidad de Keith Stanovich) recopilando test para evaluar algunas de las falacias y sesgos comentados en los capítulos precedentes.[6] Entre estos figuraban el exceso de confianza, los costos hundidos, las inconsistencias al calcular los riesgos y los efectos de la formulación (ser influido por el hecho de que un resultado se describa como una ganancia o como una pérdida). No es de extrañar que la destreza de los individuos para evitar las falacias estuviese correlacionada con su inteligencia, aunque solo parcialmente. También guardaba relación con su estilo de toma de decisiones, es decir, con el grado en que decían abordar los problemas reflexiva y constructivamente en lugar de hacerlo de manera impulsiva y fatalista.

A fin de medir las consecuencias vitales, el trío desarrolló una especie de escala de pobres desgraciados, una medida de la susceptibilidad de los individuos a los percances de toda índole. Se preguntaba a los participantes, por ejemplo, si en la última década habían estropeado prendas de vestir por

no seguir las instrucciones de lavado de la etiqueta, se habían dejado las llaves dentro del coche, habían cogido el tren o el autobús equivocado, se habían roto un hueso, habían tenido un accidente de coche, habían conducido borrachos, habían perdido dinero en acciones, se habían metido en una pelea, habían sido expulsados del colegio, habían dejado un trabajo después de una semana, o habían quedado embarazadas o dejado embarazada a alguna mujer accidentalmente. Descubrieron que las destrezas de razonamiento de las personas predecían en efecto consecuencias para sus vidas: cuantas menos falacias en el razonamiento, menos debacles en la vida.

Ni que decir tiene que correlación no es lo mismo que causalidad. La competencia en el razonamiento está correlacionada con la inteligencia en bruto, y sabemos que la inteligencia superior protege a las personas de las consecuencias negativas para su vida, como la enfermedad, los accidentes y el fracaso laboral, manteniendo constante el estatus socioeconómico.[7] Pero no es lo mismo la inteligencia que la racionalidad, pues ser buena calculando algo no es garantía de que una persona intente calcular las cosas adecuadas. La racionalidad requiere asimismo capacidad de reflexión, apertura mental y dominio de las herramientas cognitivas, como la lógica formal y la probabilidad matemática. Bruine de Bruin y sus colegas realizaron los análisis de regresión múltiple (el método explicado en el capítulo 9) y descubrieron que, incluso cuando mantenían constante la inteligencia, los mejores razonadores sufrían menos consecuencias negativas.[8]

También el estatus socioeconómico frustra nuestra fortuna en la vida. La pobreza es una carrera de obstáculos que enfrenta a las personas a los riesgos del desempleo, la toxicomanía y otras adversidades. Pero también aquí los análisis de regresión demostraron que los mejores razonadores tenían mejores resultados en la vida, manteniendo constante el estatus socioeconómico. Todo esto dista aún de demostrar la causalidad. No obstante, tenemos ya algunos de los vínculos necesarios: una alta probabilidad *a priori*, dos importantes factores de confusión estadísticamente controlados y una causalidad inversa improbable (tener un

accidente de coche no debería hacerte cometer falacias cognitivas). Esto nos autoriza a dar cierto crédito a la conclusión causal de que la competencia en el razonamiento puede proteger a una persona de los infortunios en la vida.

RACIONALIDAD Y PROGRESO MATERIAL

Aunque el sesgo de disponibilidad lo oculte a nuestra vista, el progreso humano es un hecho empírico. Cuando, más allá de los titulares, observamos las líneas de tendencia, constatamos que la humanidad es en conjunto más saludable, más rica, más longeva, está mejor alimentada, más instruida y más a salvo de las guerras, los asesinatos y los accidentes que en las décadas y los siglos pasados.[9]

Habiendo documentado estos cambios en dos libros, con frecuencia me preguntan si «creo en el progreso». La respuesta es no. Como la humorista Fran Lebowitz, no creo en nada en lo que uno tenga que creer. A pesar de que muchas medidas del bienestar humano, cuando se representan gráficamente a lo largo del tiempo, muestran un incremento gratificante (aunque no siempre ni en todas partes), este no se debe a alguna fuerza dialéctica o ley evolutiva que no cesa de elevarnos. Por el contrario, la naturaleza no tiene consideración alguna por nuestro bienestar y, con frecuencia, como en el caso de las pandemias y los desastres naturales, se diría que está tratando de destrozarnos. *Progreso* es la abreviatura de un conjunto de retrocesos y victorias cosechados en un universo implacable, y es un fenómeno que precisa explicación.

La explicación es la racionalidad. Cuando los humanos se fijan la meta de mejorar el bienestar de sus congéneres (frente a otras aspiraciones dudosas, como la gloria o la redención) y aplican su ingenio a instituciones que lo comparten con el de los demás, de vez en cuando triunfan. Cuando retienen los éxitos y toman nota de los fracasos, los beneficios pueden acumularse y designamos el panorama general como progreso.

Podemos comenzar con la más preciada de todas las cosas, la vida. A partir de la segunda mitad del siglo XIX, la esperanza de vida al nacer subió desde su histórico estancamiento en torno a los 30 años y hoy alcanza los 72,4 años a escala mundial, 83 en los países más afortunados.[10] Este don de

la vida no nos llovió del cielo. Fue el esforzado dividendo de los avances en salud pública (con el lema: «Salvar vidas, millones a la vez»), particularmente cuando la teoría microbiana de la enfermedad desplazó otras teorías causales como los miasmas, los espíritus, las conspiraciones y el castigo divino. Entre los salvavidas se incluyen la cloración del agua, los humildes retretes y el alcantarillado, el control de vectores de enfermedades como los mosquitos y las pulgas, los programas de vacunación a gran escala, la promoción del lavado de manos, y los cuidados prenatales y perinatales básicos, como la lactancia y el contacto corporal. Cuando azotan las enfermedades y las lesiones, los avances en medicina impiden que maten a tanta gente como en la época de los curanderos populares y los barberos cirujanos, incluidos los antibióticos, la antisepsia, la anestesia, las transfusiones, los medicamentos y la terapia de rehidratación oral (una solución de sal y azúcar que detiene la diarrea mortal).

La humanidad siempre ha luchado por producir suficientes calorías y proteínas para alimentarse, con la hambruna al acecho en cada mala cosecha. Pero en la actualidad se ha diezmado el hambre en la mayor parte del mundo: la desnutrición y el retraso en el crecimiento están en declive, y las hambrunas afligen hoy únicamente a las regiones más remotas y asoladas por las guerras, un problema no de escasez de alimento, sino de barreras para hacerlo llegar a los hambrientos.[11] Las calorías no llegaron del maná celestial ni en una cornucopia sostenida por Abundantia, la diosa romana de la abundancia, sino con los avances de la agronomía. Estos incluían la rotación de cultivos para reponer los suelos agotados; las tecnologías para las plantaciones y las cosechas de alto rendimiento, como las sembradoras, los arados, los tractores y las cosechadoras; los fertilizantes sintéticos (a los que se atribuye el mérito de salvar 2.700 millones de vidas); una red de transporte y almacenamiento para llevar los alimentos de la granja a la mesa, que incluye ferrocarriles, canales, camiones, graneros y refrigeración; mercados nacionales e internacionales que permiten que el excedente en una región cubra la escasez en otra; y la revolución verde de la década de 1960, que propagó los productivos y vigorosos cultivos híbridos.

La pobreza no necesita explicación; es el estado natural de la humanidad. Lo que requiere una explicación es la riqueza. Durante la mayor parte de la historia humana, en torno al 90 % de la humanidad vivía en lo que hoy designamos como pobreza extrema. En 2020, la cifra no llega al 9 %; sigue siendo demasiado alta, pero el objetivo es su erradicación en el próximo decenio.[12] El gran enriquecimiento material de la humanidad comenzó con la Revolución Industrial del siglo XIX. Esta estuvo literalmente impulsada por la captura de la energía del carbón, del petróleo, del viento y de la caída del agua, y más tarde del sol, de la Tierra y de la fisión nuclear. La energía se introducía en máquinas que transforman el calor en trabajo en las fábricas con producción en serie y en medios de transporte como ferrocarriles, canales, carreteras y buques cargueros. Las tecnologías materiales dependían de las financieras, especialmente la banca, las finanzas y los seguros. Y nada de todo ello podría haberse aprovechado para traer la prosperidad generalizada sin Gobiernos que hagan cumplir los contratos, minimicen la fuerza y el fraude, mitiguen las sacudidas financieras con bancos centrales y dinero fiable, e inviertan en bienes públicos generadores de riqueza, tales como infraestructuras, investigación básica y educación universal.

El mundo todavía no ha puesto fin a las guerras, como soñaban los cantantes folk de los sesenta, pero ha reducido drásticamente su número y su letalidad, desde una cifra de 21,9 muertes en combate por cien mil habitantes en 1950 hasta solo 0,7 en 2019.[13] El trío Peter, Paul y Mary merece algún reconocimiento, pero mayor es el mérito de las instituciones diseñadas para reducir los incentivos de las naciones para ir a la guerra, comenzando con el proyecto de una «paz perpetua» de Immanuel Kant en 1795. Una de ellas es la democracia, que, como vimos en el capítulo sobre correlación y causalidad, reduce realmente las probabilidades de guerra, presumiblemente porque la carne de cañón de un país es menos aficionada a ese pasatiempo que sus reyes y generales. Otra es el comercio y las inversiones internacionales, que hacen más barato comprar las cosas que robarlas, y vuelve imprudente que los países maten a sus clientes y deudores (la Unión Europea, galardonada con el Premio Nobel de la Paz en 2012, nació de una organización comercial, la Comunidad Europea del

Carbón y del Acero). Otra es una red de organizaciones internacionales, particularmente Naciones Unidas, que une a los países en una comunidad, moviliza fuerzas por el mantenimiento de la paz, inmortaliza Estados, exime de fronteras, y proscribe y estigmatiza la guerra al tiempo que proporciona medios alternativos para resolver las disputas.

Las creaciones del ingenio humano han garantizado asimismo otros impulsos al bienestar, tales como la seguridad, el ocio, los viajes y el acceso al arte y al entretenimiento. Aunque muchos de los artilugios y las burocracias crecieron orgánicamente y fueron perfeccionados mediante ensayo y error, ninguno de ellos fue un accidente. En su momento se defendían con argumentos impulsados por la lógica y las evidencias, las causas y los efectos, así como los equilibrios entre los intereses individuales y el bien común. Tendremos que redoblar nuestro ingenio para abordar las pruebas a las que hoy nos enfrentamos, en particular la tragedia de los comunes del carbono (capítulo 8). Tendremos que aplicar nuestras capacidades intelectuales a las tecnologías que abaraten las energías limpias, a la tarificación que encarezca las energías sucias, a las políticas que impidan que las facciones se conviertan en expoliadoras, y a los tratados que hagan globales y equitativos los sacrificios.[14]

RACIONALIDAD Y PROGRESO MORAL

El progreso consiste en algo más que en avances en la seguridad y el bienestar material. Consiste asimismo en avances en la forma de tratarnos unos a otros, en igualdad, benevolencia y derechos. Muchas prácticas crueles e injustas se han reducido en el transcurso de la historia. Entre estas figuran los sacrificios humanos, la esclavitud, el despotismo, los deportes sangrientos, el eunuquismo, los harenes, el vendado de pies, los sádicos castigos corporales y capitales, la persecución de herejes y disidentes, y la opresión de las mujeres y de las minorías religiosas, raciales, étnicas y sexuales.[15] Ninguna de ellas ha sido extirpada de la faz de la Tierra, pero cuando representamos gráficamente los cambios históricos, en todos los casos vemos descensos y, en algunos, desplomes.

¿Cómo hemos llegado a disfrutar de este progreso? Theodore Parker y, un siglo más tarde, Martin Luther King Jr., intuyeron un arco moral que se curvaba hacia la justicia. Pero la naturaleza del arco y su capacidad de tirar de las palancas del comportamiento humano son misteriosas. Cabe imaginar sendas más prosaicas: modas cambiantes, campañas de denuncia, apelaciones al corazón, movimientos populares de protesta, cruzadas religiosas y moralizadoras. Una idea popular es que el progreso moral se logra mediante la lucha: los poderosos jamás renuncian a sus privilegios, que han de serles arrebatados por el poder de la gente que actúa de forma solidaria.[16]

Mi mayor sorpresa al intentar comprender el progreso moral es cuántas veces en la historia la primera ficha del dominó fue un argumento razonado.[17] Un filósofo redactaba un escrito que argumentaba por qué alguna práctica era indefendible, irracional o incompatible con los valores que todos afirmaban defender. El panfleto o manifiesto se hacía viral, se traducía a otras lenguas, se debatía en pubs, salones y cafés, y acababa influyendo en los dirigentes, los legisladores y la opinión popular. Finalmente, la conclusión era absorbida por la sabiduría convencional y la decencia común de una sociedad, borrando el rastro de los argumentos que la introdujeron en ella. Hoy en día pocas personas sienten la necesidad o poseen la capacidad de formular un argumento coherente sobre los motivos por los que son malos la esclavitud, el destripamiento o las palizas a los niños; resulta evidente. Sin embargo, exactamente esos debates tuvieron lugar hace siglos.

Y los argumentos que prevalecieron, cuando se llama nuestra atención sobre ellos en la actualidad, continúan sonando convincentes. Apelan a un sentido de la razón que trasciende los siglos, porque concuerdan con los principios de coherencia conceptual que forman parte de la realidad misma. Ahora bien, como vimos en el capítulo 2, ningún argumento lógico puede establecer una afirmación moral. Pero un argumento *puede* establecer que una afirmación que es objeto de debate es inconsistente con otra afirmación profundamente valorada por alguien, o con valores como la vida y la felicidad que la mayoría de las personas reclaman para sí mismas y convendrían en que son deseos legítimos de todo el mundo. Como vimos en

el capítulo 3, la inconsistencia es fatal para el razonamiento: un conjunto de creencias que incluye una contradicción puede desplegarse para deducir cualquier cosa y es perfectamente inútil.

Como he de ser receloso a la hora de inferir la causación de la correlación y de señalar un único motivo en una malla de entrecruzamientos históricos, no puedo afirmar que los buenos argumentos sean la causa del progreso moral. No podemos realizar un ensayo controlado aleatorio sobre la historia, en la que a la mitad de la muestra de las sociedades se la exponga a un tratado moral convincente y a la otra mitad se le administre un placebo lleno de galimatías moralistas. Tampoco contamos con un conjunto de datos sobre triunfos morales suficientemente grande como para extraer una conclusión causal de la red de correlaciones (lo más cercano que se me ocurre son los estudios transnacionales que demuestran que la educación y el acceso a la información en una época, que son indicadores de la voluntad de intercambiar ideas, predicen la democracia y los valores liberales en una era posterior, manteniendo constantes los factores de confusión socioeconómicos).[18] Por el momento solo puedo ofrecer ejemplos de argumentos precoces, que los historiadores nos cuentan que fueron influyentes en su día y que siguen siendo impecables en nuestro tiempo.

COMENCEMOS CON LA PERSECUCIÓN RELIGIOSA. ¿De verdad necesitaba la gente un argumento intelectual para entender por qué podía ser un poquito malo eso de quemar herejes en la hoguera? Lo cierto es que sí. En 1553, el teólogo francés Sebastián Castellio (1515-1563) elaboró un argumento contra la intolerancia religiosa, advirtiendo la ausencia de razonamiento detrás de las ortodoxias de Juan Calvino y la «consecuencia lógica» de sus prácticas:

> Calvino dice que está seguro, y [otras sectas] también dicen estarlo; Calvino dice que están equivocadas y desea juzgarlos, y ellos lo mismo. ¿Quién habrá de ser el juez? ¿Quién nombró a Calvino árbitro de todas las sectas, para que solo él matara? Él tiene la Palabra de Dios y ellos también. Si la cuestión es segura, ¿para quién lo es? ¿Para Calvino? Pero entonces, ¿por qué escribe tantos libros acerca de la verdad manifiesta? [...] En vista de la incertidumbre, hemos de definir al hereje simplemente como alguien de quien discrepamos. Y entonces, si vamos a matar herejes, la consecuencia lógica será una guerra de exterminio, toda vez que cada uno está seguro

> de sí mismo. Calvino tendría que invadir Francia y todas las demás naciones, arrasar ciudades, pasar a espada a todos los habitantes, sin perdonar a ningún sexo ni a ninguna edad, ni siquiera a los bebés ni a los animales.[19]

El siglo XVI conoció otro argumento precoz en contra de una práctica bárbara. Hoy en día parece evidente que la guerra no es saludable para los niños y demás seres vivientes. Sin embargo, durante la mayor parte de la historia, la guerra se consideraba noble, santa, emocionante, varonil y gloriosa.[20] Aunque la guerra solo dejó de ser venerada después de los cataclismos del siglo XX, las semillas del pacifismo habían sido sembradas por uno de los «padres de la modernidad», el filósofo Erasmo de Róterdam (1466-1536), en su ensayo de 1517 *Alegato de la razón, la religión y la humanidad contra la guerra*. Tras ofrecer una descripción conmovedora de los beneficios de la paz y los horrores de la guerra, Erasmo pasa a un análisis de la guerra en clave de decisión racional, explicando sus resultados de suma cero y su utilidad esperada negativa:

> Añádase a estas consideraciones que las ventajas derivadas de la paz se difunden por doquier, y alcanzan a un gran número, mientras que en la guerra, si algo acaba felizmente [...], la ventaja redunda solo en unos pocos, que son indignos de cosecharla. La seguridad de un hombre se debe a la destrucción de otro; la recompensa de un hombre deriva del saqueo de otro. La causa de regocijo para un bando es para el otro una causa de duelo. Todo lo desdichado en la guerra lo es ciertamente con gravedad y, por el contrario, todo lo que se denomina buena fortuna es una buena fortuna salvaje y cruel, una felicidad mezquina, que deriva su existencia del infortunio ajeno. De hecho, la conclusión suele ser que ambos bandos, el victorioso y el vencido, tienen motivos para lamentarse. No sé de ninguna guerra que haya trascurrido tan felizmente en todos sus acontecimientos que el conquistador, si tenía un corazón para sentir o un entendimiento para juzgar, como debería ser el caso, no se arrepintiera de haber participado en ella [...].
>
> Si tuviéramos que calcular con justicia el asunto, y hacer un cómputo justo del coste de ocuparse de la guerra y de procurar la paz, descubriríamos que la paz podría comprarse con una décima parte de los cuidados, trabajos, problemas, peligros, gastos y sangre que cuesta llevar a cabo una guerra [...].
>
> Pero el objetivo es infligir todo el daño posible al enemigo. ¡Un objetivo sumamente inhumano! Y consideremos si podemos dañarle de forma contundente sin dañar al mismo tiempo, y por los mismos medios, a nuestra propia gente. Sin duda es una locura llevarte una parte tan grande de un mal seguro cuando siempre ha de ser incierto cómo caerá al final el dado de la guerra.[21]

La Ilustración del siglo XVIII fue una fuente de argumentos contra otras clases de crueldad y opresión. Como sucede con la persecución religiosa, nos quedamos casi sin palabras cuando nos preguntan qué tiene de malo el

uso de la tortura sádica en los castigos penales, como el arrastre y descuartizamiento, la ruptura de articulaciones en la rueda, la quema en la hoguera o el aserrado por la mitad desde la entrepierna. Sin embargo, en un panfleto de 1764, el economista y filósofo utilitarista Cesare Beccaria (1738-1794) expuso argumentos en contra de tales barbaries identificando los costes y los beneficios de la pena criminal. El objetivo legítimo del castigo, argüía Beccaria, consiste en incentivar a los individuos para que no exploten a los demás, y la utilidad esperada de la maldad debería ser la métrica mediante la cual evaluar las prácticas punitivas.

> A medida que los castigos se vuelven más crueles, las mentes de los hombres, que, como los fluidos, siempre se ajustan al nivel de su entorno, se endurecen, y el poder siempre vivaz de las emociones hace que, después de cien años de torturas crueles, la rueda no cause más temor del que antes producía la prisión. Para que un castigo sirva a su propósito, es suficiente con que el daño que inflija supere el beneficio que el criminal pueda obtener del crimen, y en el cálculo de este balance hemos de añadir la certeza del castigo y la pérdida del bien producido por el crimen. Cualquier cosa más allá de esto resulta superflua y, por ende, tiránica.[22]

El argumento de Beccaria, y de sus colegas *philosophes* Voltaire y Montesquieu, influyó en la prohibición de «los castigos crueles e inusuales» por la octava enmienda de la Constitución estadounidense. En los últimos años se sigue invocando la enmienda para socavar el repertorio de ejecuciones en Estados Unidos, y muchos observadores legales creen que la declaración de inconstitucionalidad de todas estas prácticas es solo cuestión de tiempo.[23]

Otras formas de barbarie se pusieron en el punto de mira durante la Ilustración mediante argumentos que siguen siendo penetrantes hasta el día de hoy. El otro gran utilitarista del siglo XVIII, Jeremy Bentham (1748-1832), elaboró el primer argumento sistemático en contra de la criminalización de la homosexualidad:

> En cuanto a cualquier daño principal, es evidente que no produce dolor a nadie. Por el contrario, produce placer [...]. Las dos personas lo desean. Si cualquiera de ellas no lo desea, el acto no es lo que estamos considerando aquí; es un delito totalmente diferente en la naturaleza de sus efectos: es un daño personal; es una clase de violación [...]. En cuanto a cualquier peligro distinto del dolor, el peligro, si existe, ha de consistir en la tendencia del ejemplo. Ahora bien, ¿cuál es la tendencia de este ejemplo? Disponer a otros a llevar a cabo las mismas prácticas: pero esta práctica, por lo que conocemos hasta ahora, no provoca a nadie ninguna clase de dolor.[24]

Bentham formuló asimismo el argumento en contra de la crueldad hacia los animales de un modo que continúa orientando en nuestros días el movimiento de protección de los animales:

> *Puede* que llegue el día en que el resto de la creación animal adquiera esos derechos que jamás podrían habérseles denegado más que de la mano de la tiranía. Los franceses ya han descubierto que la negrura de la piel no es razón para que se abandone sin reparación a un ser humano al capricho de un verdugo. Puede que llegue el día en que se reconozca que el número de patas, la vellosidad de la piel o la terminación del hueso sacro [coxis] son razones igualmente insuficientes para abandonar a un ser sintiente a la misma suerte. ¿Qué otra cosa debería trazar la línea insuperable? ¿Acaso la facultad de la razón o tal vez la facultad del discurso? Pero un caballo o un perro adulto es, más allá de cualquier comparación, un animal más racional y más sociable que un bebé de un día, una semana o incluso un mes. Pero supongamos que no fuese así. ¿Qué cambiaría? La cuestión no es si pueden *razonar* ni si pueden *hablar*, sino si pueden *sufrir*.[25]

La yuxtaposición de Bentham de las diferencias moralmente irrelevantes en el color de piel entre los humanos con las diferencias en los rasgos físicos y cognitivos entre las especies no es un mero símil. Es un acicate para cuestionar nuestra respuesta instintiva a las características superficiales de los entes que nos piden que consideremos (la reacción del sistema 1, si se quiere) y para reflexionar sobre la coherencia de nuestras creencias acerca de quién es merecedor de derechos y protecciones.

Estimular la reflexión cognitiva comparando un grupo protegido con otro vulnerable es un medio común del que se han valido los persuasores morales para que la gente cobre conciencia de sus sesgos e intolerancias. El filósofo Peter Singer, un descendiente intelectual de Bentham y el más destacado defensor actual de los derechos de los animales, designa el proceso como *círculo en expansión*.[26]

La esclavitud fue un marco de referencia habitual. La Ilustración albergó un vigoroso movimiento abolicionista, iniciado por los argumentos de Jean Bodin (1530-1596), John Locke (1632-1704) y Montesquieu (1689-1755).[27] En el caso de los dos últimos, su denuncia de la esclavitud fue asimismo la base de su crítica de la monarquía absoluta y de su insistencia en que el poder legítimo de los Gobiernos solo dimana del consentimiento de los gobernados. El punto de partida era socavar el supuesto de una jerarquía natural: cualquier clasificación de aristócrata y plebeyo, señor y

vasallo, amo y esclavo. «Nacemos libres —escribía Locke— por cuanto nacemos racionales.»[28] Los humanos somos seres intrínsecamente pensantes, sintientes y volitivos, y nadie posee un derecho natural a dominar a cualquier otro. En su capítulo sobre la esclavitud en *Two Treatises of Government (Dos tratados sobre el Gobierno civil)*, Locke explica:

> La libertad de los hombres bajo el Gobierno consiste en la posesión de una regla de conducta vigente, común para todos los hombres de esa sociedad y cuya elaboración ha corrido a cargo del poder legislativo erigido en ella; una libertad que me permite seguir mi propia voluntad en todo aquello que la norma no prescriba, así como no estar sometido a la voluntad inconstante, incierta, desconocida y arbitraria de otro hombre: siendo una libertad natural, no ha de someterse a más restricción que la ley de la naturaleza.[29]

La idea clave de que la igualdad es la relación por defecto entre las personas fue incorporada por Thomas Jefferson (1743-1826) como justificación del gobierno democrático: «Consideramos evidentes en sí mismas estas verdades: que todos los hombres son creados iguales, que están dotados por su Creador de ciertos derechos inalienables; que entre estos están la vida, la libertad y la búsqueda de la felicidad; que, para garantizar estos derechos, se instituyen entre los hombres los gobiernos, que derivan sus poderes legítimos del consentimiento de los gobernados».

Aunque Locke pudo haber previsto que sus escritos inspirarían uno de los grandes desarrollos de la historia humana, el advenimiento de la democracia, puede que no previera otro de los que inspiró. En su prefacio de 1730 a *Some Reflections upon Marriage [Algunas reflexiones sobre el matrimonio]*, la filósofa Mary Astell (1666-1731) escribió:

> Si la soberanía absoluta no es necesaria en un Estado, ¿cómo llega a serlo en una familia? O, si lo es en una familia, ¿por qué no en un Estado? Pues no cabe alegar ninguna razón para la una que no se sostenga con más firmeza para la otra [...]. Si todos los hombres nacen libres, ¿por qué todas las mujeres nacen esclavas? ¿No han de serlo, si el estar sujetas a la voluntad inconstante, incierta, desconocida y arbitraria de los hombres es la perfecta condición de la esclavitud?[30]

¿Suena familiar? Astell se apropió astutamente del argumento de Locke (incluida su frase «la perfecta condición de la esclavitud») a fin de socavar la opresión de las mujeres, lo que la convierte en la primera feminista inglesa. Mucho antes de que llegase a ser un movimiento

organizado, el feminismo comenzó siendo un argumento, recogido después de Astell por la filósofa Mary Wollstonecraft (1759-1797). En su libro de 1792 *A Vindication of the Rights of Women (Vindicación de los derechos de la mujer)*, Wollstonecraft no solo extendió el argumento de que era lógicamente inconsistente negar a las mujeres los derechos garantizados a los hombres, sino que adujo asimismo que el supuesto de que las mujeres poseen un intelecto o una autoridad inherentemente inferiores a los de los hombres era espurio, debido a una confusión entre naturaleza y crianza: las mujeres eran criadas sin la educación ni las oportunidades otorgadas a los hombres. Comenzaba su libro con una carta abierta a Talleyrand, figura destacada de la Revolución francesa, que había sostenido que, con perdón de la *égalité*, las chicas no necesitaban una educación formal:

> Considere, me dirijo a usted como legislador, si, cuando los hombres luchan por su igualdad, y para que se les permita juzgar por sí mismos, respetando su propia felicidad, no resulta incoherente e injusto subyugar a las mujeres, aun cuando uno crea firmemente estar obrando de la manera mejor calculada para promover su felicidad. ¿Quién hizo al hombre el juez exclusivo, si la mujer participa con él en el don de la razón?
>
> De esta guisa argumentan los tiranos de toda denominación, desde el débil rey hasta el débil padre de familia; todos ellos están deseosos de aplastar la razón; sin embargo, todos afirman que usurpan su trono solamente en aras de la utilidad. ¿No interpreta usted un papel similar cuando obliga a todas las mujeres, negándoles los derechos civiles y políticos, a permanecer encerradas en sus familias tanteando en la oscuridad? Pues sin duda, señor, no afirmará usted que un deber no fundado en la razón pueda ser vinculante. Si ese es en efecto su destino, los argumentos pueden extraerse de la razón; y, con tan augusto respaldo, cuanto más entendimiento adquieran las mujeres, más apegadas estarán a su deber, comprendiéndolo, pues a menos que lo comprendan, a menos que su moralidad esté basada en los mismos principios inmutables que los del hombre, ninguna autoridad puede hacerles cumplir con él de una manera virtuosa. Pueden ser esclavas útiles, pero la esclavitud ejercerá su efecto permanente, degradando al amo y a la esclava sumisa.[31]

Y hablando de la esclavitud misma, los argumentos verdaderamente imponentes en contra de la abominable institución llegaron de la mano del escritor, editor y estadista Frederick Douglass (1818-1895). Nacido él mismo en la esclavitud, Douglass podía despertar la intensa empatía de su público por el sufrimiento de los esclavos y, como uno de los más grandes oradores de la historia, podía conmover con la música y las imágenes de su discurso. No obstante, Douglass desplegaba estos dones al servicio de la argumentación moral rigurosa. En su más célebre discurso de 1852, «What

to the slave is the Fourth of July?» [«¿Qué es el 4 de julio para el esclavo?»], Douglass rechazaba apofáticamente cualquier necesidad de ofrecer argumentos contra la esclavitud empleando «las reglas de la lógica», porque, decía, eran evidentes, para proceder acto seguido a hacer exactamente eso. Por ejemplo:

> Hay setenta y dos crímenes en el estado de Virginia que, si son cometidos por un hombre negro, por muy ignorante que sea, le someten a la pena de muerte, mientras que solo dos de los mismos crímenes someterán al mismo castigo a un hombre blanco. ¿Qué es esto, sino el reconocimiento de que el esclavo es un ser moral, inteligente y responsable? Se concede la humanidad del esclavo. Esta se admite en el hecho de que los códigos sureños están repletos de leyes que prohíben, bajo severas multas y sanciones, enseñar a leer o a escribir al esclavo. Cuando podáis mostrarme alguna de esas leyes en referencia a las bestias del campo, puedo consentir debatir la humanidad del esclavo.[32]

Douglass proseguía: «Y en un momento como este se necesita ironía abrasadora, no argumentos convincentes», y luego se enfrentaba a su público con un copioso inventario de inconsistencias en sus sistemas de creencias:

> Lanzáis vuestros anatemas a los tiranos de cabeza coronada de Rusia y Austria, y os enorgullecéis de vuestras instituciones democráticas, mientras vosotros mismos consentís ser los meros instrumentos y guardaespaldas de los tiranos de Virginia y Carolina. Invitáis a vuestras costas a los fugitivos de la opresión extranjera, los honráis con banquetes, los recibís con ovaciones, los aclamáis, brindáis por ellos, los homenajeáis, los protegéis y los inundáis con vuestro dinero como si de agua se tratase: pero a los fugitivos de vuestro propio país los anunciáis, cazáis, arrestáis, disparáis y matáis [...].
>
> Podéis descubrir vuestro pecho ante la tormenta de la artillería británica para libraros de un impuesto de tres peniques sobre el té; y, sin embargo, exprimís hasta el último cuarto de penique de los trabajadores negros de vuestro país.

Y presagiando a Martin Luther King más de un siglo después, recordaba a la nación su declaración fundacional:

> Declaráis ante el mundo, y el mundo comprende que declaráis, que consideráis «evidentes en sí mismas estas verdades: que todos los hombres son creados iguales, que están dotados por su Creador de ciertos derechos inalienables; y que entre estos están la vida, la libertad y la búsqueda de la felicidad»; y, sin embargo, mantenéis sujeta con firmeza, en una esclavitud que, de acuerdo con vuestro propio Thomas Jefferson, «es peor que los tiempos contra los que vuestros padres se alzaron en rebelión», a una séptima parte de los habitantes de vuestro país.

El hecho de que Douglass y King pudieran citar con aprobación a Jefferson, él mismo un hombre hipócrita y en ciertos sentidos deshonroso, no compromete la racionalidad de sus respectivos argumentos, sino que la refuerza. Deberíamos preocuparnos de la virtud de las personas cuando las consideramos amigas, pero no cuando consideramos las ideas que expresan. Las ideas son verdaderas o falsas, consistentes o contradictorias, conducentes o no al bienestar humano, con independencia de quien las piense. La igualdad de los seres sintientes, fundamentada en la irrelevancia lógica de la distinción entre *yo* y *tú*, es una idea que las personas a través de los tiempos redescubren, transmiten y hacen extensiva a nuevos seres vivos, expandiendo el círculo de la compasión como una energía oscura moral.

Los argumentos sólidos, que imponen la coherencia de nuestras prácticas con nuestros principios y con la meta del florecimiento humano, no pueden mejorar el mundo por sí mismos. Pero han guiado y deberían guiar los movimientos transformadores. Marcan la diferencia entre la fuerza moral y la fuerza bruta, entre las marchas por la justicia y las turbas de linchamiento, entre el progreso humano y la destrucción. Y serán argumentos sólidos tanto para revelar plagas morales como para descubrir remedios viables, lo que necesitaremos para garantizar que continúe el progreso moral, que las prácticas abominables de hoy lleguen a resultarles tan increíbles a nuestros descendientes como lo son para nosotros la quema de herejes y la subasta de esclavos.

El poder de la racionalidad para guiar el progreso moral va de la mano de su poder para orientar el progreso material y las decisiones sabias en nuestras vidas. Nuestra capacidad para lograr incrementar el bienestar en un cosmos implacable y para ser buenos con los demás a pesar de nuestra naturaleza imperfecta depende de que captemos principios imparciales que trasciendan nuestra experiencia provinciana. Somos una especie que ha sido dotada de una facultad elemental de razonar y que ha descubierto fórmulas e instituciones que amplían su alcance. Estas nos despiertan a ideas y nos exponen a realidades desconcertantes para nuestras intuiciones, pero que son verdaderas a pesar de todo.

Referencias bibliográficas

Abelson, R. P., «Beliefs are like possessions», *Journal for the Theory of Social Behaviour*, 16, 1986, págs. 223-250, <https://doi.org/10.1111/j.1468-5914.1986.tb00078.x>.

Abito, J. M., y Salant, Y., «The effect of product misperception on economic outcomes: Evidence from the extended warranty market», *Review of Economic Studies*, 86, 2018, págs. 2285-2318, <https://doi.org/10.1093/restud/rdy045>.

Acerbi, A., «Cognitive attraction and online misinformation», *Palgrave Communications*, 5, 2019, págs. 1-7, <https://doi.org/10.1057/s41599019-0224-y>.

Aggarwal, C. C., *Neural Networks and Deep Learning*, Nueva York, Springer, 2018.

Ainslie, G., *Breakdown of Will*, Nueva York, Cambridge University Press, 2001.

Alexander, S., «Conflict vs. mistake», *Slate Star Codex*, 2018, <https://slatestarcodex.com/2018/01/24/conflictvsmistake/>.

Ali, R., *Dear Colleague Letter*, US Department of Education, 2011 (orientación sobre políticas del subsecretario de Derechos Civiles), <https://www2ed.gov/about/offices/list/ocr/letters/colleague-201104.html>.

Allais, M., «Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école americaine», *Econometrica*, 21, 1953, págs. 503-546, <https://doi.org/10.2307/1907921>.

American Academy of Arts and Sciences, *Perceptions of Science in America*, Cambridge (MA), American Academy of Arts and Sciences, 2018.

Appiah, K. A., *The Honor Code: How Moral Revolutions Happen*, Nueva York, W. W. Norton, 2010.

Arbital, «Bayes' rule», 2020, <https://arbital.com/p/bayes_rule/?l=1zq>.

Arkes, H. R., Gigerenzer, G., y Hertwig, R., «How bad is incoherence?», *Decision*, 3, 2016, págs. 20-39, <https://doi.org/10.1037/dec0000043>.

Arkes, H. R., y Mellers, B. A., «Do juries meet our expectations?», *Law and Human Behavior*, 26, 2002, págs. 625-639, <https://doi.org/10.1023/A:1020929517312>.

Armstrong, S. L., Gleitman, L. R., y Gleitman, H., «What some concepts might not be», *Cognition*, 13, 1983, págs. 263-308, <https://doi.org/10.1016/0010-0277(83)90012-4>.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., y Waldron, E. M., «A neuropsychological theory of multiple systems in category learning», *Psychological Review*, 105, 1998, págs. 442-481, <https://doi.org/10.1037/0033-295X.105.3.442>.

Astell, M., *Some Reflections upon Marriage: To Which Is Added a Preface, in Answer to Some Objections*, Farmington Hills (MI), Gale ECCO, 2010 [1730].

Bacon, F., *Novum organum*, Seattle, Washington, CreateSpace, 2017 [1620] (trad. cast.: *La gran restauración (Novum organum)*, Madrid, Tecnos, 2011).

Bankoff, C., «Dick Cheney simply does not care that the CIA tortured innocent people», *The New York Magazine*, 14 de diciembre de 2014, <https://nymag.com/intelligencer/2014/12/cheney-alright-with-torture-of-innocent-people.html>.

Bar-Hillel, M., «The base-rate fallacy in probability judgments», *Acta Psychologica*, 44, 1980, págs. 211-233, <https://doi.org/10.1016/00016918(80)90046-3>.

Baron, J., «Applying evidence to social programs», *The New York Times*, 29 de noviembre de 2012, <https://economix.blogs.nytimes.com/2012/11/29/applying-evidence-to-social-programs/>.

—, «Actively open-minded thinking in politics», *Cognition*, 188, 2019, págs. 8-18, <https://doi.org/10.1016/j.cognition.2018.10.004>.

—, y Jost, J. T., «False equivalence: Are liberals and conservatives in the United States equally biased?», *Perspectives on Psychological Science*, 14, 2019, págs. 292-303, <https://doi.org/10.1177/1745691618788876>.

Basterfield, C., Lilienfeld, S. O., Bowes, S. M., y Costello, T. H., «The Nobel disease: when intelligence fails to protect against irrationality», *Skeptical Inquirer*, mayo de 2020, <https://skepticalinquirer.org/2020/05/thenobel-disease-when-intelligence-fails-to-protect-against-irrationality/>.

Batt, J., *Stolen Innocence: A Mother's Fight for Justice – The Authorized Story of Sally Clard*, Londres, Ebury Press, 2004.

Baumard, N., y Boyer, P., «Religious beliefs as reflective elaborations on intuitions: A modified dualprocess model», *Current Directions in Psychological Science*, 22, 2013, págs. 295-300, <https://doi.org/10.1177/0963721413478610>.

Baumeister, R. F., Stillwell, A., y Wotman, S. R., «Victim and perpetrator accounts of interpersonal conflict: Autibiographical narratives about anger», *Journal of Personality and Social Psychology*, 59, 1990, págs. 994-1005, <https://doi.org/10.1037/0022-3514.59.5.994>.

Baumeister, R. F., y Tierney, J., *Willpower: Rediscovering the Greatest Human Strength*, Londres, Penguin, 2012.

Bazelon, E., y Larimore, R., «How often do women falsely cry rape?», *Slate*, 1 de octubre de 2009, <https://slate.com/news-and-politics/2009/10/why-it-s-so-hard-to-quantify-false-rape-charges.html>.

BBC, «Avoid gold teeth, says Turkmen leader», *BBC News*, 7 de abril de 2004, <http://news.bbc.co.uk/2/hi/asiapacific/3607467.stm>.

—, «The Crown: Netflix has "no plans" for a fiction warning», *BBC News*, 6 de diciembre de 2020, <https://www.bbc.com/news/entertainmentarts-55207871>.

Beccaria, C., *On Crimes and Punishments and Other Writings*, Nueva York, Cambridge University Press, 2010 [1764] (trad. cast.: *De los delitos y las penas*, Madrid, Biblioteca Nueva, 2018).

Bell, E. T., *The Development of Mathematics*, Nueva York, McGraw Hill, 1947 (2.ª ed.) (trad. cast.: *Historia de las matemáticas*, México, FCE, 1949).

Bentham, J., *An Introduction to the Principles of Morals and Legislation*, 1789, <https://www.econlib.org/library/Bentham/bnthPML.html> (trad. cast.: *Introducción a los principios de la moral y la legislación*, Barcelona, Marbot, 2015).

—, y Crompton, L., «Offences against one's self: Paederasty (part I)», *Journal of Homosexuality*, 3, 1978 [1785], págs. 389-405, <https://doi.org/10.1300/J082v03n04_07>.

Benthem, J. van, «Logic and reasoning: Do the facts matter?», *Studia Logica*, 88, 2008, págs. 67-84, <https://doi.org/10.1007/s11225-008-9101-1>.

Binmore, K., *Fun and Games: A Text on Game Theory*, Boston, Houghton Mifflin, 1991 (trad. cast.: *Teoría de juegos*, Madrid, McGraw-Hill, 1996).

—, *Game Theory: A Very Short Introduction*, Nueva York, Oxford University Press, 2007 (trad. cast.: *La teoría de juegos: una breve introducción*, Madrid, Alianza, 2011).

—, «Do conventions need to be common knowledge?», *Topoi*, 27, 2008, págs. 17-27, <https://doi-org.ezp-prod1.hul.harvard.edu/10.1007/s11245-008-9033-4>.

Blackwell, M., «Black Lives Matter and the mechanics of conformity», *Quillette*, 17 de septiembre de 2020, <https://quillettte.com/2020/09/17/black-lives-matter-and-the-mechanics-of-conformity/>.

Block, W., *Defending the Undefendable*, Auburn (AL), Ludwig von Mises Institute, 2018 [1976] (trad. cast.: *Defendiendo lo indefendible*, Madrid, Unión Editorial, 2012).

Bloom, P., *Descartes' Baby: How the Science of Child Development Explains What Makes Us Human*, Nueva York, Basic Books, 2004.

Bond, M., «Risk school», *Nature*, 461, págs. 1189-1192, 28 de octubre, 2009.

Bornstein, D., «The dawn of the evidence-based budget», *The New York Times*, 30 de mayo de 2012, <https://opinionator.blogs.nytimes.com/2012/05/30/worthy-of-government-funding-prove-it>.

—, y Rosenberg, T., «When reportage turns to cynicism», *The New York Times*, 14 de noviembre de 2016, <https://www.nytimes.com/2016/11/15/opinion/when-reportage-turns-to-cynicism.html>.

Braverman, B., «Why you should steer clear of extended warranties», *Consumer Reports*, 22 de diciembre de 2018, <https://www.consumerreports.org/extended-warranties/steer-clear-extended-warranties/>.

Breyer, S., *Breaking the Vicious Circle: Toward Effective Risk Regulation*, Cambridge (MA), Harvard University Press, 1993.

Bruine de Bruin, W., Parker, A. M., y Fischhoff, B., «Individual differences in adult decision-making competence», *Journal of Personality and Social Psychology*, 92, 2007, págs. 938-956, <https://doi.org/10.1037/00223514.92.5.938>.

Brunvand, J. H., *Too Good to Be True: The Colossal Book of Urban Legends*, Nueva York, W. W. Norton, 2014 (trad. cast.: *El fabuloso libro de las leyendas urbanas*, Barcelona, Debolsillo, 2008).

Bump, P., «Trump's effort to steal the election comes down to some utterly ridiculous statistical claims», *The Washington Post*, 9 de diciembre de 2020, <https://www.washingtonpost.com/politics/2020/12/09/trumps-effort-steal-election-comes-down-some-utterly-ridiculous-statistical-claims/>.

Burns, K., «At veterinary colleges, male students are in the minority», *American Veterinary Medical Association*, 15 de febrero de 2010, <https://www.avma.org/javma-news/2010-02-15/veterinary-colleges-male-students-are-minority>.

Caldeira, K., Emanuel, K., Hansen, J., y Wigley, T., «Top climate change scientists' letter to policy influencers», CNN, 3 de noviembre de 2013, <https://www.cnn.com/2013/11/03/world/nuclear-energy-climatechange-scientists-letter/index.html>.

Campbell, B., y Manning, J., *The Rise of Victimhood Culture: Microaggressions, Safe Spaces, and the New Culture Wars*, Londres, Palgrave Macmillan, 2018.
Caplan, B., «What's wrong with the rationality community», *EconLog*, 4 de abril de 2017, <https://www.econlib.org/archives/2017/04/whats_wrong_wit_22.html>.
Carroll, L., «What the tortoise said to Achilles», *Mind*, 4, 1895, págs. 178-180.
—, «Symbolic logic», en W. W. Bartley (comp.), *Lewis Carroll's Symbolic Logic*, Nueva York, Clarkson Potter, 1977 [1896].
Carroll, S. M., *The Big Picture: On the Origins of Life, Meaning, and the Universe Itself*, Nueva York, Penguin Random House, 2016.
Cesario, J., y Johnson, D. J., «Statement on the retraction of "Officer characteristics and racial disparities in fatal officerinvolved shootings"», 2020, <https://doi.org/10.31234/osf.io/dj57k>.
Chagnon, N. A., *Yanomamö*, Fort Worth, Texas, Harcourt Brace, 1997 (5.ª ed.).
Chapman, L. J., y Chapman, J. P., «Genesis of popular but erroneous psychodiagnostic observations», *Journal of Abnormal Psychology*, 72, 1967, págs. 193-204, <https://doi.org/10.1037/h0024670>.
—, «Illusory correlation as an obstacle to the use of valid psychodiagnostic signs», *Journal of Abnormal Psychology*, 74, 1969, págs. 271-280, <https://doi.org/10.1037/h0027592>.
Charlesworth, T. E. S., y Banaji, M. R., «Patterns of implicit and explicit attitudes: I. Long-term change and stability from 2007 to 2016», *Psychological Science*, 30, 2019, págs. 174-192, <https://doi.org/10.1177/0956797618813087>.
Cheng, P. W., y Holyoak, K. J., «Pragmatic reasoning schemas», *Cognitive Psychology*, 17, 1985, págs. 391-416, <https://doi.org/10.1016/00100285(85)90014-3>.
Chivers, T., *The AI Does Not Hate You: Superintelligence, Rationality and the Race to Save the World*, Londres, Weidenfeld & Nicolson, 2019.

Chomsky, N., *Language and Mind*, Nueva York, Cambridge University Press, 2006 [1972] (trad. cast.: *El lenguaje y el entendimiento*, Barcelona, Seix Barral, 1986).

Chwe, M. S.-Y., *Rational Ritual: Culture, Coordination, and Common Knowledge*, Princeton (NJ), Princeton University Press, 2001.

Clegg, L. F., «Protean free will», manuscrito inédito, California Institute of Technology, 2012, <https://resolver.caltech.edu/CaltechAUTHORS:20120328-152031480>.

Cohen, I. B., *Science and the Founding Fathers: Science in the Political Thought of Thomas Jefferson, Benjamin Franklin, John Adams, and James Madison*, Nueva York, W. W. Norton, 1997.

Cohn, A., Maréchal, M. A., Tannenbaum, D., y Zünd, C. L., «Civic honesty around the globe», *Science*, 365, 2019, págs. 70-73, <https://doi.org/10.1126/science.aau8712>.

Cohon, R., «Hume's moral philosophy», en E. N. Zalta (comp.), *The Stanford Encyclopedia of Philosophy*, 2018, <https://plato.stanford.edu/entries/hume-moral/>.

Cole, M., Gay, J., Glick, J., y Sharp, D. W., *The Cultural Context of Learning and Thinking*, Nueva York, Basic Books, 1971.

Combs, B., y Slovic, P., «Newspaper coverage of causes of death», *Journalism Quarterly*, 56, 1979, págs. 837-849.

Cosmides, L., «The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task», *Cognition*, 31, 1989, págs. 187-276, <https://doi.org/10.1016/0010-0277(89)90023-1>.

—, y Tooby, J., «Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty», *Cognition*, 58, 1996, págs. 1-73, <https://doi.org/10.1016/00100277(95)00664-8>.

Coyne, J. A., *Faith versus Fact: Why Science and Religion Are Incompatible*, Nueva York, Penguin, 2015.

Crockett, Z., «The time everyone "corrected" the world's smartest woman», *Priceonomics*, 19 de febrero de 2015, <https://priceonomics.com/thetime-everyone-corrected-the-worlds-smartest/>.

Curtis, G. N., «The Fallacy Files taxonomy of logical fallacies», 2020, <https://www.fallacyfiles.org/taxonnew.htm>.

Dasgupta, P., «The Stern Review's economics of climate change», *National Institute Economic Review*, 199, 2007, págs. 4-7, <https://doi.org/10.1177/0027950107077111>.

Davis, D. B., *Slavery and Human Progress*, Nueva York, Oxford University Press, 1984.

Dawes, R. M., Faust, D., y Meehl, P. E., «Clinical versus actuarial judgment», *Science*, 243, 1989, págs. 1668-1674, <https://doi.org/10.1126/science.2648573>.

Dawkins, R., *The Selfish Gene*, Nueva York, Oxford University Press, 2016 [1976] (trad. cast.: *El gen egoísta: las bases biológicas de nuestra conducta*, Barcelona, Salvat, 1994).

—, *The God Delusion*, Nueva York, Houghton Mifflin, 2006 (trad. cast.: *El espejismo de Dios*, Barcelona, Espasa, 2017).

Dawson, E., Gilovich, T., y Regan, D. T., «Motivated reasoning and performance on the Wason selection task», *Personality and Social Psychology Bulletin*, 28, 2002, págs. 1379-1387, <https://doi.org/10.1177/014616702236869>.

Deary, I. J., *Intelligence: A Very Short Introduction*, Nueva York, Oxford University Press, 2001 (trad. cast.: *Una brevísima introducción a la inteligencia*, México, Océano, 2004).

DellaVigna, S., y Kaplan, E., «The Fox News effect: Media bias and voting», *Quarterly Journal of Economics*, 122, 2007, págs. 1187-1234, <https://doi.org/10.1162/qjec.122.3.1187>.

Dennett, D. C., *Breaking the Spell: Religion As a Natural Phenomenon*, Nueva York, Penguin, 2006 (trad. cast.: *Romper el hechizo: la religión como un fenómeno natural*, Madrid, Katz, 2011).

—, *Intuition Pumps and Other Tools for Thinking*, Nueva York, W. W. Norton, 2013 (trad. cast.: *Bombas de intuición y otras herramientas de pensamiento*, México, FCE, 2015).

Ditto, P. H., Clark, C. J., Liu, B. S., *et al*., «Partisan bias and its discontents», *Perspectives on Psychological Science*, 14, 2019, págs. 304-316, <https://doi.org/10.1177/1745691618817753>.

Ditto, P. H., Liu, B. S., Clark, C. J., *et al*., «At least bias is bipartisan: A meta-analytic comparison of partisan bias in liberals and conservatives», *Perspectives on Psychological Science*, 14, 2019, págs. 273-291, <https://doi.org/10.1177/1745691617746796>.

Donaldson, H., Doubleday, R., Hefferman, S., *et al*., «Are talking heads blowing hot air? An analysis of the accuracy of forecasts in the political arena», Hamilton College, 2011, <https://www.hamilton.edu/documents/Analysis-of-Forecast-Accuracy-in-the-Political-Media.pdf>.

Douglass, F., «What to the slave is the Fourth of July?», en P. S. Foner (comp.), *Frederick Douglass: Selected Speeches and Writings*, Chicago, Lawrence Hill, 1999 [1852].

Dukas, H., y Hoffman, B. (comps.), *Albert Einstein, The Human Side: New Glimpses from his Archives*, Princeton (NJ), Princeton University Press, 1981.

Duffy, B., *The Perils of Perception: Why We're Wrong about Nearly Everything*, Londres, Atlantic Books, 2018.

Eagle, A., «Chance versus randomness», en E. N. Zalta (comp.), *The Stanford Encyclopedia of Philosophy*, 2019, <https://plato.stanford.edu/entries/chance-randomness/>.

Earman, J., «Bayes, Hume, Price, and miracles», *Proceedings of the British Academy*, 113, 2002, págs. 91-109.

Edwards, A. W. F., «Is the Pope an alien?», *Nature*, 382, 1996, pág. 202, <https://doi.org/10.1038/382202b0>.

Eisenstein, E. L., *The Printing Revolution in Early Modern Europe*, Nueva York, Cambridge University Press, 2012 (2.ª ed.) (trad. cast.: *La revolución de la imprenta en la Edad Moderna europea*, Madrid, Akal, 1994).

Eliot, G., *Essays of George Eliot*, Filadelfia, Routledge, 2017 [1883].
Ellickson, R. C., *Order without Law: How Neighbors Settle Disputes*, Cambridge (MA), Harvard University Press, 1991.
Elster, J. (comp.), *Deliberative Democracy*, Nueva York, Cambridge University Press, 1998 (trad. cast.: *La democracia deliberativa*, Barcelona, Gedisa, 2001).
Emerson, R. W., *Self-reliance and Other Essays*, Nueva York, Dover, 1993 [1841] (trad. cast.: *Confianza en uno mismo*, Madrid, Gadir, 2009).
Erasmus [Erasmo de Róterdam], D., *The Complaint of Peace: To Which Is Added, Antipolemus; Or, the Plea of Reason, Religion, and Humanity, against War*, Miami, Florida, HardPress, 2017 [1517] (trad. cast.: *Lamento de la paz*, Barcelona, Acantilado, 2020).
Erceg, N., Galić, Z., y Bubić, A., «"Dysrationalia" among university students: The role of cognitive abilities, different aspects of rational thought and self-control in explaining epistemically suspect beliefs», *Europe's Journal of Psychology*, 15, 2019, págs. 159-175, <https://doi.org/10.5964/ejop.v15i1.1696>.
Evans, J. S. B. T., «Dual-process theories of deductive reasoning: Facts and fallacies», en K. J. Holyoak y R. G. Morrison (comps.), *The Oxford Handbook of Thinking and Reasoning*, Oxford, Oxford University Press, 2012.
Fabrikant, G., «Humbler, after a streak of magic», *The New York Times*, 11 de mayo de 2008, <https://www.nytimes.com/2008/05/11/business/11bill.html>.
FBI, «Crime in the United States, expanded homicide data table 1», 2019, <https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s-2019/tables/expanded-homcide-data-table-1.xls>.
Federal Aviation Administration, *Pilot's Handbook of Aeronautical Knowledge*, Oklahoma City, U. S. Department of Transportation, 2016, <https://www.faa.gov/regulations_policies/handbooks_manuals/aviation/phak/media/pilot_handbook.pdf>.
Feller, W., *An Introduction to Probability Theory and Its Applications*, Nueva York, Wiley, 1968 (trad. cast.: *Introducción a la teoría de probabilidades y sus aplicaciones*, México, Limusa, 1978).

Fiddick, L, Cosmides, L., y Tooby, J., «No interpretation without representation: The role of domainspecific representations and inferences in the Wason selection task», *Cognition*, 77, 2000, págs. 1-79, <https://doi.org/10.1016/S0010-0277(00)00085-8>.

Finkel, E. J., Bail, C. A., Cikara, M., *et al*., «Political sectarianism in America», *Science*, 370, 2020, págs. 533-536, <https://doi.org/10.1126/science.abe1715>.

Fishkin, J. S., *When the People Speak: Deliberative Democracy and Public Consultation*, Nueva York, Oxford University Press, 2011.

Flaherty, C., «Failure to communicate: Professor suspended for saying a Chinese word that sounds like a racial slur in English», *Inside Higher Ed*., 2020, <https://www.insidehighered.com/news/2020/09/08/professor-suspended-saying-chinese-word-sounds-english-slur>.

Freitas, J. de, Thomas, K., DeScioli, P., y Pinker, S., «Common knowledge, coordination, and strategic mentalizing in human social life», *Proceedings of the National Academy of Sciences*, 116, 2019, págs. 13751-13758, <https://doi.org/10.1073/pnas.1905518116>.

Fodor, J. A., *Psychological Explanation: An Introduction to the Philosophy of Psychology*, Nueva York, Random House, 1968 (trad. cast: *La explicación psicológica*, Madrid, Cátedra, 1991).

Fox, C., «Social media: How might it be regulated?», *BBC News*, 12 de noviembre de 2020, <https://www.bbc.com/news/technology-54901083>.

Frank, R. H., *Passions within Reason: The Strategic Role of the Emotions*, Nueva York, W. W. Norton, 1988.

Frederick, S., «Cognitive reflection and decision making», *Journal of Economic Perspectives*, 19, 2005, págs. 25-42, <https://doi.org/10.1257/089533005775196732>.

French, C., «Precognition studies and the curse of the failed replications», *The Guardian*, 15 de marzo de 2012, <http://www.theguardian.com/science/2012/mar/15/precognition-studies-curse-failedreplications>.

Friedersdorf, C., «Why can't people hear what Jordan Peterson is actually saying?», *The Atlantic*, 22 de enero de 2018, <https://www.theatlantic.com/politics/archive/2018/01/putting-monsterpaint-onjordan-peterson/550859/>.

Friesen, J. P., Campbell, T. H., y Kay, A. C., «The psychological advantage of unfalsibiability: The appeal of untestable religious and political ideologies», *Journal of Personality and Social Psychology*, 108, 2015, págs. 515-529, <https://doi.org/10.1037/pspp0000018>.

Galton, F., «Regression toward mediocrity in hereditary stature», *Journal of the Anthropological Institute of Great Britain and Ireland*, 15, 1886, págs. 246-263.

Gampa, A., Wokcik, S. P., Motyl, M., *et al*., «(Ideo)logical reasoning: Ideology impairs sound reasoning», *Social Psychological and Personality Science*, 10, 2019, págs.1075-1083, <https://doi.org/10.1177/1948550619829059>.

Gardner, M., «Problems involving questions of probability and ambiguity», *Scientific American*, 201, 1959, págs. 174-182.

—, «Why the long arm of coincidence is usually not as long as it seems», *Scientific American*, 227, 1972.

Gelman, A., y Loken, E., «The statistical crisis in science», *American Scientist*, 102, 2014, págs. 460-465.

Gelman, S. A., *The Essential Child: Origins of Essentialism in Everyday Thought*, Nueva York, Oxford University Press, 2005.

Gettier, E. L., «Is justified true belief knowledge?», *Analysis*, 23, 1963, págs. 121-123.

Gigerenzer, G., «How to make cognitive illusions disappear: Beyond "heuristics and biases"», *European Review of Social Psychology*, 2, 1991, págs. 83-115, <https://doi.org/10.1080/14792779143000033>.

—, «On narrow norms and vague heuristics: A reply to Kahneman and Tversky», *Psychological Review*, 103, 1996, págs. 592-596, <https://doi.org/10.1037/0033-295X.103.3.592>.

—, «Ecological intelligence: An adaptation for frequencies», en D. D. Cummins y C. Allen (comps.), *The Evolution of Mind*, Nueva York, Oxford University Press, 1998.

—, «Gigerenzer's Law of indispensable ignorance», *Edge*, 2004, <https://www.edge.org/response-detail/10224>.

—, «Out of the frying pan into the fire: Behavioral reactions to terrorist attacks», *Risk Analysis*, 26, 2006, págs. 347-351, <https://doi.org/10.1111/j.1539-6924.2006.00753.x>.

—, «The evolution of statistical thinking», en G. Gigerenzer (comp.), *Rationality for Mortals: How People Cope with Uncertainty*, Nueva York, Oxford University Press, 2008.

—, *Rationality for Mortals: How People Cope with Uncertainty*, Nueva York, Oxford University Press, 2008.

—, «What are natural frequencies?», *BMJ*, 343, 2011, d6386, <https://doi.org/10.1136/bmj.d6386>.

—, «Breast cancer screening pamphlets mislead women», *BMJ*, 348, 2014, g2636, <https://doi.org/10.1136/bmj.g2636>.

—, «On the supposed evidence for libertarian paternalism», *Review of Philosophy and Psychology*, 6, 2015, págs. 361-383, <https://doi.org/10.1007/s13164-015-0248-1>.

—, «The Bias Bias in behavioral economics», *Review of Behavioral Economics*, 5, 2018, págs. 303-336, <https://doi.org/10.1561/105.00000092>.

—, «Statistical rituals: The replication delusion and how we got there», *Advances in Methods and Practices in Psychological Science*, 1, 2018, págs. 198-218, <https://doi.org/10.1177/2515245918771329>.

—, Hertwig, R., Broek, E. van den, Fasolo, B., y Katsikopoulos, K. V., «"A 30 % chance of rain tomorrow": How does the public understand probabilistic weather forecasts?», *Risk Analysis: An International Journal*, 25, 2005, págs. 623-629, <https://doi.org/10.1111/j.1539-6924.2005.00608.x>.

—, Krauss, S., y Vitouch, O., «The null ritual: What you always wanted to know about significance testing but were afraid to ask», en D. Kaplan (comp.), *The Sage Handbook of Quantitative Methodology for the Social Sciences*, Thousand Oaks (CA), Sage, 2004.

—, Swijtink, Z., Porter, T., *et al*., *The Empire of Chance: How Probability Changed Science and Everyday Life*, Nueva York, Cambridge University Press, 1989.

—, y García-Retamero, R., «Cassandra's regret: The psychology of not wanting to know», *Psychological Review*, 124, 2017, págs. 179-196.

—, y Kolpatzik, K., «How new fact boxes are explaining medical risk to millions», *BMJ*, 357, 2017, j2460, <https://10.1136/bmj.j2460>.

Gilbert, B., «The 10 most-viewed fake-news stories on Facebook in 2019 were just revealed in a new report», *Business Insider*, 6 de noviembre de 2019, <https://www.businessinsider.com/most-viewed-fake-newsstories-shared-on-facebook-2019-2019-11>.

Gilovich, T., Vallone, R., y Tversky, A., «The hot hand in basketball: On the misperception of random sequences», *Cognitive Psychology*, 17, 1985, págs. 295-314, <https://doi.org/10.1016/0010-0285(85)90010-6>.

Glaeser, E. L., «Psychology and the market», *The American Economic Review*, 94, 2004, págs. 408-413, <http://www.jstor.org/stable/3592919>.

Goda, G. S., Levy, M. R., Manchester, C. F., *et al*., «The role of time preferences and exponential-growth bias in retirement savings», *National Bureau of Economic Research Working Paper Series*, 21482, 2015, <https://doi.org/10.3386//w21482>.

Goldstein, J. S., «Chicken dilemmas: crossing the road to cooperation», en I. W. Zartman y S. Touval (comps.), *International Cooperation: The Extents and Limits of Multilateralism*, Nueva York, Cambridge University Press, 2010.

—, *Winning the War on War: The Decline of Armed Conflict Worldwide*, Nueva York, Penguin, 2011.

—, Qvist, S. A., y Pinker, S., «Nuclear power can save the world», *The New York Times*, 6 de abril de 2019, <https://www.nytimes.com/2019/04/06/opinion/sunday/climate-change-nuclear-power.html>.

—, y Qvist, S. A., *A Bright Future: How Some Countries Have Solved Climate Change and the Rest Can Follow*, Nueva York, PublicAffairs, 2019.

Goldstein, R. N., *Betraying Spinoza: The Renegade Jew Who Gave Us Modernity*, Nueva York, Nextbook/Shocken, 2006.

—, *36 Arguments for the Existence of God: A Work of Fiction*, Nueva York, Pantheon, 2010.

—, *Plato at the Googleplex: Why Philosophy Won't Go Away*, Nueva York, Pantheon, 2013.

Goldstein-Rose, S., *The 100 % Solution: A Plan for Solving Climate Change*, Nueva York, Melville House, 2020.

Good, I., «When batterer becomes murderer», *Nature*, 381, 1996, pág. 481, <https://doi.org/10.1038/381481a0>.

Goodfellow, I., Bengio, Y., y Courville, A., *Deep Learning*, Cambridge (MA), MIT Press, 2016.

Gould, S. J., «The streak of streaks», *The New York Review of Books*, 1988, <https://www.nybooks.com/articles/1988/08/18/the-streak-ofstreaks/>.

—, *Rocks of Ages: Science and Religion in the Fullness of Life*, Nueva York, Ballantine, 1999 (trad. cast.: *Ciencia versus religión: un falso conflicto*, Barcelona, Booket, 2012).

Gracyk, T., «Hume's aesthetics», en E. N. Zalta (comp.), *Stanford Encyclopedia of Philosophy*, 2020, <https://plato.stanford.edu/archives/sum2020/entries/hume-aesthetics/>.

Granberg, D., y Brown, T. A., «The Monty Hall dilemma», *Personality & Social Psychology Bulletin*, 21, 1995, págs. 711-723, <https://doi.org/10.1177/0146167295217006>.

Grayling, A. C., *Toward the Light of Liberty: The Struggles for Freedom and Rights that Made the Modern Western World*, Nueva York, Walker, 2007.

Green, D. M., y Swets, J. A., *Signal Detection Theory and Psychophysics*, Nueva York, Wiley, 1966.

Greene, J., *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*, Nueva York, Penguin, 2013.

Grice, H. P., «Logic and conversation», en P. Cole y J. L. Morgan (comps.), *Syntax & Semantics*, vol. 3, *Speech Acts*, Nueva York, Academic Press, 1975.

Haidt, J., *The Righteous Mind: Why Good People Are Divided by Politics and Religion*, Nueva York, Pantheon, 2012 (trad. cast.: *La mente de los justos: por qué la política y la religión dividen a la gente sensata*, Barcelona, Deusto, 2019).

—, «Why universities must choose one telos: truth or social justice», *Heterodox Academy*, 16 de octubre de 2016, <https://heterodoxacademy.org/blog/one-telos-truth-or-social-justice-2/>.

Hájek, A., «Interpretations of probability», en E. N. Zalta (comp.), *The Stanford Encylopedia of Philosophy*, 2019, <https://plato.stanford.edu/archives/fall2019/entries/probability-interpret/>.

Hallsworth, M., y Kirkman, E., *Behavioral Insights*, Cambridge (MA), MIT Press, 2020.

Hamilton, I. A., «Jeff Bezos explains why his best decisions were based off intuition, not analysis», *Inc.*, 14 de septiembre de 2018, <https://www.inc.com/business-insider/amazon-ceo-jeff-bezos-says-his-bestdecision-were-made-when-he-followed-his-gut.html>.

Harris, S., *The End of Faith: Religion, Terror, and the Future of Reason*, Nueva York, W. W. Norton, 2005 (trad. cast.: *El fin de la fe: la religión, el terror y el futuro de la razón*, Madrid, Paradigma, 2007).

Hastie, R., y Dawes, R. M., *Rational Choice in an Uncertain World: The Psychology of Judgment and Decision Making*, Los Ángeles, Sage, 2010 (2.ª ed.).

Henderson, L., «The problem of induction», en E. N. Zalta (comp.), *The Stanford Encyclopedia of Philosophy*, 2020, <https://plato.stanford.edu/archives/spr2020/entries/induction-problem>.

Henrich, J., Heine, S. J., y Norenzayan, A., «The weirdest people in the world?», *Behavioral and Brain Sciences*, 33, 2010, págs. 61-83, <https://doi.org/10.1017/S0140525X0999152X>.

Hertwig, R., y Engel, C., «Homo ignorans: Deliberately choosing not to know», *Perspectives on Psychological Science*, 11, 2016, págs. 359-372.

Hertwig, R., y Gigerenzer, G., «The “conjuction fallacy” revisited: How intelligent inferences look like reasoning errors», *Journal of Behavioral Decision Making*, 12, 1999, págs. 275-305, <https://doi.org/10.1002/(SICI)1099-0771(199912)12:4<275::AID-BDM323>3.0.CO;2-M>.

Hobbes, T., *Leviathan*, Nueva York, Oxford University Press, 1957 [1651] (trad. cast.: *Leviatán, o la materia, forma y poder de un Estado eclesiástico y civil*, Madrid, Alianza, 2014).

Hoffrage, U., Lindsey, S., Hertwig, R., y Gigerenzer, G., «Communicating statistical information», *Science*, 290, 2000, págs. 2261-2262, <https://doi.org/10.1126/science.290.5500.2261>.

Holland, P. W., «Statistics and causal inference», *Journal of the American Statistical Association*, 81, 1986, págs. 945-960, <https://doi.org/10.2307/2289064>.

Homer [Homero], *The Odyssey*, Nueva York, Norton, 2018 [700 a. C.] (trad. cast.: *Odisea*, Barcelona, Nueva Biblioteca Clásica Gredos, 2019).

Hood, B., *Supersense: Why we Believe in the Unbelievable*, Nueva York, HarperCollins, 2009.

Horowitz, D. L., *The Deadly Ethnic Riot*, Berkeley, University of California Press, 2001.

Hume, D., *A Treatise of Human Nature*, Nueva York, Oxford University Press, 2000 [1739] (trad. cast.: *Tratado de la naturaleza humana: autobiografía*, Madrid, Tecnos, 2002).

—, *An Enquiry Concerning Human Understanding*, Nueva York, Oxford University Press, 1999 [1748] (trad. cast.: *Investigación sobre el entendimiento humano*, Madrid, Istmo, 2004).

Hunt, L., *Inventing Human Rights: A History*, Nueva York, W. W. Norton, 2007 (trad. cast.: *La invención de los derechos humanos*, Barcelona, Tusquets, 2009).

Ichikawa, J. J., y Steup, M., «The analysis of knowledge», en E. N. Zalta (comp.), *The Stanford Encyclopedia of Philosophy*, 2018, <https://plato.stanford.edu/entries/knowledge-analysis/>.

Ioannidis, J. P. A., «Why most published research findings are false», *PloS Medicine*, 2, 2005, e124, <https://doi.org/10.1371/journal.pmed.0020124>.

James, W., *The Principles of Psychology*, Nueva York, Dover, 1950 [1890] (trad. cast.: *Principios de psicología*, México, FCE, 1989).

Jarvis, S., Deschenes, O., y Jha, A., *The Private and External Costs of Germany's Nuclear Phase-out*, 2019, <https://haas.berkeley.edu/wpcontent/uploads/WP304.pdf>.

Jenkins, S., «The Crown's fake history is as corrosive as fake news», *The Guardian*, 16 de noviembre de 2020, <http://www.theguardian.com/commentisfree/2020/nov/16/the-crown-fake-history-news-tv-seriesroyal-family-artistic-licence>.

Jeszeck, C. A., Collins, M. J., Glickman, M., Hoffrey, L., y Grover, S., «Retirement security: Most households approaching retirement have low savings», *United States Government Accountability Office*, 2015, <https://www.gao.gov/assets/680/670153.pdf>.

Johnson, D. J., y Cesario, J., «Reply to Knox and Mummolo and Schimmack and Carlsson: Controlling for crime and population rates», *Proceedings of the National Academy of Sciences*, 117, 2020, págs. 1264-1265, <https://doi.org/10.1073/pnas.1920184117>.

Johnson, D. J., Tress, T., Burkel, N., *et al.*, «Officer characteristics and racial disparities in fatal officerinvolved shootings», *Proceedings of the National Academy of Sciences*, 116, 2019, págs. 15877-15882, <https://doi.org/10.1073/pnas.1903856116>.

Johnson, S., *The Letters of Samuel Johnson with Mrs. Thrale's Genuine Letters to Him*, Nueva York, Oxford University Press, 1963.

Jones, J. M., «Confidence in higher education down since 2015», *Gallup Blog*, 9 de octubre de 2018, <https://news.gallup.com/opinion/gallup/242441/confidencehigher-educationdown2015.aspx>.

Joyner, J., «Ranking the pundits: A study shows that most national columnists and talking heads are about as accurate as a coin flip», *Outside the Beltway*, 3 de mayo de 2011, <https://www.outsidethebeltway.com/ranking-the-pundits/>.

Kaba, M., «Yes, we mean literally abolish the police», *The New York Times*, 12 de junio de 2020, <https://www.nytimes.com/2020/06/12/opinion/sunday/floydabolish-defundpolice.html>.

Kahan, D. M., «Ideology, motivated reasoning, and cognitive reflection», *Judgment and Decision Making*, 8, 2013, págs. 407-424, <http://dx.doi.org/10.2139/ssrn.2182588>.

—, «Climate-science communication and the measurement problem», *Political Psychology*, 36, 2015, págs. 1-43, <https://doi.org/10.1111/pops.12244>.

Kahan, D. M., Hoffman, D. A., Braman, D., *et al*., «“They saw a protest”: Cognitive illiberalism and the speech-conduct distinction», *Stanford Law Review*, 64, 2012, págs. 851-906.

Kahan, D. M., Peters, E., Dawson, E. C., y Slovic, P., «Motivated numeracy and enlightened self-goverment», *Behavioural Public Policy*, 1, 2017, págs. 54-86, <10.1017/bpp.2016.2>.

Kahan, D. M., Peters, E., Wittlin, M., *et al*., «The polarizing impact of science litteracy and numeracy on perceived climate change risks», *Nature Climate Change*, 2, 2012, págs. 732-735, <https://doi.org/10.1038/nclimate1547>.

Kahan, D. M., Wittlin, M., Peters, E., *et al*., «The tragedy of the risk-perception commons: Culture conflict, rationality conflict, and climate change», *Yale Law & Economics Research Paper*, 435, 2011, <http://dx.doi.org/10.2139/ssrn.1871503>.

Kahneman, D., «Daniel Kahneman-facts», The Nobel Prize, 2002, <https://www.nobelprize.org/prizes/economic-sciences/2002/kahneman/facts/>.

—, *Thinking, Fast and Slow*, Nueva York, Farrar, Straus and Giroux, 2011 (trad. cast.: *Pensar rápido, pensar despacio*, Barcelona, Debate, 2012).

—, Slovic, P., y Tversky, A., *Judgment Under Uncertainty: Heuristics and Biases*, Nueva York, Cambridge University Press, 1982.

Kahneman, D., y Tversky, A., «Subjective probability: A judgment of representativeness», *Cognitive Psychology*, 3, 1972, 3, págs. 430-454, <https://doi.org/10.1016/0010-0285(72)90016-3>.

—, «Prospect theory: An analysis of decisions under risk», *Econometrica*, 47, 1979, págs. 313-327, <https://doi.org/10.1142/9789814417358_0006>.

—, «On the reality of cognitive illusions», *Psychological Review*, 103, 1996, págs. 582-591, <https://doi.org/10.1037/0033-295X.103.3.582>.

Kaplan, R. D., «The Coming Anarchy», *The Atlantic*, 1994, <https://www.theatlantic.com/magazine/archive/1994/02/the-coming-anarchy/304670/>.

Kelemen, D., y Rosset, E., «The human function compunction: teleological explanation in adults», *Cognition*, 111, 2009, págs. 138-143, <https://doi.org/10.1016/j.cognition.2009.01.001>.

Kendler, K. S., Kessler, R. C., Walters, E. E., *et al*., «Stressful life events, genetic liability, and onset of an episode of major depression in women», *Focus*, 8, 2010, págs. 459-470, <https://doi.org/10.1176/foc.8.3.foc.459>.

Kenny, C., *Getting better: Why Global Development Is Succeeding – And How We Can Improve the World even More*, Nueva York, Basic Books, 2011.

Kessler, G., Rizzo, S., y Kelly, M., «Trump is averaging more than 50 false or misleading claims a day», *The Washington Post*, 22 de octubre de 2020, <https://www.washingtonpost.com/politics/2020/10/22/presidenttrump-is-averaging-more-than-50-false-or-misleading-claims-day/>.

King, G., Keohane, R. O., y Verba, S., *Designing Social Inquiry: Scientific Inference in Qualitative Research*, Princeton (NJ), Princeton University Press, 1994 (trad. cast.: *El diseño de la investigación social: la inferencia científica en los estudios cualitativos*, Madrid, Alianza, 2018).

Kingdon, J., *Self-Made Man: Human Evolution from Eden to Extinction?*, Nueva York, Wiley, 1993.

Kissinger, H., «How the Enlightenment ends», *The Atlantic*, junio de 2018, <https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/>.

Knox, D., y Mummolo, J., «Making inferences about racial disparities in police violence», *Proceedings of the National Academy of Sciences*, 117, 2020, págs. 1261-1262, <https://doi.org/10.1073/pnas.1919418117>.

Kors, A. C., y Silverglate, H. A., *The Shadow University: The Betrayal of Liberty on America's Campuses*, Nueva York, Free Press, 1998.

Kräenbring, J., Monzon Penza, T., Gutmann, J., *et al*., «Accuracy and completeness of drug information in Wikipedia: A comparison with standard textbooks of pharmacology», *PloS ONE*, 9, 2014, e106930, <https://doi.org/10.1371/journal.pone.0106930>.

Krämer, W., y Gigerenzer, G., «How to confuse with statistics or: the use and misuse of conditional probabilities», *Statistical Science*, 20, 2005, págs. 223-230, <https://doi.org/10.1214/088342305000000029>.

Kunda, Z., «The case for motivated reasoning», *Psychological Bulletin*, 108, 1990, págs. 480-498, <https://doi.org/10.1037/0033-2909.108.3.480>.

Laibson, D., «Golden eggs and hyperbolic discounting», *Quarterly Journal of Economics*, 112, 1997, págs. 443-477, <https://doi.org/10.1162/003355397555253>.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., y Gershman, S. J., «Building machines that learn and think like people», *Behavioral and Brain Sciences*, 39, 2017, págs. 1-101, <https://doi.org/10.1017/S0140525X16001837>.

Lane, R., «A truth reckoning: Why we're holding those who lied for Trump accountable», *Forbes*, 7 de enero de 2021, <https://www.forbes.com/sites/randalllane/2021/01/07/a-truth-reckoning-why-were-holding-those-who-lied-for-trump-accountable/?sh=5fedd2605710>.

Lankford, A., y Madfis, E., «Don't name them, don't show them, but report everything else: A pragmatic proposal for denying mass killers the attention they seek and deterring future offenders», *American Behavioral Scientist*, 62, 2018, págs. 260-279, <https://doi.org/10.1177/0002764217730854>.

Lazari-Radek, K. de, y Singer, P., «The objectivity of ethics and the unity of practical reason», *Ethics*, 123, 2012, págs. 9-31, <https://doi.org/10.1086/667837>.

Lee, B., y Daly, R. (comps.), *The Cambridge Encyclopedia of Hunters and Gatherers*, Cambridge (RU), Cambridge University Press, 1999.

Lehrer, J., «The truth wears off», *The New Yorker*, 5 de diciembre de 2010, <https://www.newyorker.com/magazine/2010/12/13/the-truth-wears-off>.

Leibniz, G. W., «On universal synthesis and analysis, or the art of discovery and judgment», en L. E. Loemker (comp.), *Philosophical Papers and Letters*, Nueva York, Springer, 1989 [1679].

Levitt, S. D., y Dubner, S. J., *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*, Nueva York, William Morrow, 2009 (trad. cast.: *Freakonomics: Un economista políticamente incorrecto explora el lado oculto de lo que nos afecta*, Barcelona, B de Bolsillo, 2018.

Lewis, D. K., *Convention: A Philosophical Study*, Cambridge (MA), Harvard University Press, 1969.

Lewis, M., *The Undoing Project: A Friendship That Changed Our Minds*, Nueva York, Norton, 2016 (trad. cast.: *Deshaciendo errores: Kahneman, Tversky y la amistad que nos enseñó cómo funciona la mente*, Barcelona, Debate, 2017).

Lewis, M. K., *Too Dumb to Fail: How the GOP Betrayed the Reagan Revolution to Win Elections (and How It Can Reclaim Its Conservative Roots*), Nueva York, Hachette, 2016.

Lewis-Kraus, G., «The great AI awakening», *The New York Times Magazine*, pág. 12, 14 de diciembre de 2016, <https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html>.

Liberman, M. Y., «If P, so why not Q?», *Language Log*, 5 de agosto de 2004, <http://itre.cis.upenn.edu/~myl/languagelog/archives/001314.html>.

Lichtenstein, S., y Slovic, P., «Reversals of preference between bids and choices in gambling decisions», *Journal of Experimental Psychology*, 89, 1971, págs. 46-55, <https://doi.org/10.1037/h0031207>.

Liebenberg, L., *The Art of Tracking: The Origin of Science*, Cape Town, David Philip, 1990.

—, «Notes on tracking and trapping: Examples of hunter-gatherer ingenuity», manuscrito inédito, 2020, <https://stevenpinker.com/files/pinker/files/liebenberg.pdf>.

—, *The Origin of Science: The Evolutionary Roots of Scientific Reasoning and its Implications for Tracking Science*, Cape Town, CyberTracker, 2021 [2013], <https://cybertracker.org/downloads/tracking/Liebenberg-2013-The-Origin-of-Science.pdf>.

—, Am//Ao y Lombard, M., «Tracking Science: An Alternative for Those Excluded by Citizen Science», *Citizen Science Theory and Practice*, 6(1), 2021.

Lilienfeld, S. O., Ammirati, R., y Landfield, K., «Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare?», *Perspectives on Psychological Science*, 4, 2009, págs. 390-398, <https://doi.org/10.1111/j.1745-6924.2009.01144.x>.

Locke, J., *The Second Treatise of Civil Government*, Peterborough, Broadview Press, 2015 [1689] (trad. cast.: *Segundo tratado sobre el Gobierno civil*, Madrid, Alianza, 2014).

Lockwood, A. H., Welker-Hood, K., Rauch, M., y Gottlieb, B., *Coal's Assault on Human Health: A Report from Physicians for Social Responsibility*, <https://www.psr.org/blog/resource/coals-assault-on-humanhealth/>, 2009.

Loftus, E. F., Doyle, J. M., Dysart, J. E., y Newirth, K. A., *Eyewitness Testimony: Civil and Criminal*, Dayton (OH), LexisNexis, 2019 (6.ª ed.).

Lord, C. G., Ross, L., y Lepper, M. R., «Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence», *Journal of Personality and Social Psychology*, 37, 1979, págs. 2098-2109, <https://doi.org/10.1037/0022-3514.37.11.2098>.

Luce, R. D., y Raiffa, H., *Games and Decisions: Introduction and Critical Survey*, Nueva York, Dover, 1957.

Lukianoff, G., *Unlearning liberty: Campus Censorship and the End of American Debate*, Nueva York, Encounter Books, 2012.

—, y Haidt, J., *The Coddling of the American Mind: How Good Intentions and Bad Ideas Are Setting Up a Generation for Failure*, Nueva York, Penguin, 2018.

Lynn, S. K., Wormwood, J. B., Barrett, L. F., y Quigley, K. S., «Decision making from economic and signal detection perspectives: Development of an integrated framework», *Frontiers in Psychology*, 6, 2015, <https://doi.org/10.3389/fpsyg.2015.00952>.

Lyttleton, J., «Social media is determined to slow the spread of conspiracy theories like QAnon. Can they?», *Millennial Source*, 28 de octubre de 2020, <https://themilsource.com/2020/10/28/social-media-determined-to-slow-spread-conspiracy-theories-like-qanon-can-they/>.

MacAskill, W., *Doing Good Better: Effective Altruism and How You Can Make a Difference*, Nueva York, Penguin, 2015.

Maines, R., «Why are women crowding into schools of veterinary medicine but are not lining up to become engineers?», *Cornell Chronicle*, 12 de junio de 2007, <https://news.cornell.edu/stories/2007/06/why-womenbecome-veterinarians-not-engineers>.

Mann, T. E., y Ornstein, N. J., *It's Even Worse Than It Looks: How the American Constitutional System Collided with the New Politics of Extremism*, Nueva York, Basic Books, 2016 [2012].

Marcus, G. F., «Two kinds of representation», en E. Dietrich y A. B. Markman (comps.), *Cognitive Dynamics: Conceptual and Representational Change in Humans and Machines*, Mahwah (NJ), Erlbaum, 2000.

—, «The deepest problem with deep learning», *Medium*, 1 de diciembre de 2018, <https://medium.com/@GaryMarcus/the-deepest-problem-with-deep-learning-91c5991f5695>.

—, y Davis, E., *Rebooting AI: Building Artificial Intelligence We Can Trust*, Nueva York, Penguin Random House, 2019.

Marlowe, F., *The Hadza: Hunter-gatherers of Tanzania*, Berkeley, University of California Press, 2010.

Martin, G. J., y Yurukoglu, A., «Bias in cable news: Persuasion and polarization», *American Economic Review*, 107, 2017, págs. 2565-2599, <https://doi.org/10.1257/aer.20160812>.

Maymin, P. Z., y Langer, E. J., «Cognitive biases and mindfulness», *Humanities and Social Sciences Communications*, 8, pág. 40, 2021, <https://doi.org/10.1057/s41599-021-00712-1>.

Maynard Smith, J., *Evolution and the Theory of Games*, Nueva York, Cambridge University Press, 1982.

McCarthy, J., «More Americans say crime is rising in U. S.», *Gallup*, 22 de octubre de 2015, <https://news.gallup.com/poll/186308/americanssay-crime-rising.aspx>.

—, «Americans still greatly overestimate U.S. gay population», *Gallup*, 2019, <https://news.gallup.com/poll/259571/americans-greatly-over-estimate-gay-population.aspx>.

McCawley, J. D., *Everything That Linguists Have Always Wanted to Know about Logic, but Were Ashamed to Ask*, Chicago, University of Chicago Press, 1993 (2.ª ed.).

McClure, S. M., Laibson, D., Loewenstein, G., y Cohen, J. D., «Separate neural systems value immediate and delayed monetary rewards», *Science*, 306, 2004, págs. 503-507, <https://doi.org/10.1126/science.1100907>.

McGinn, C., *Truth by Analysis: Games, Names, and Philosophy*, Nueva York, Oxford University Press, 2012.

McNeil, B. J., Pauker, S. G., Sox, H. C., Jr., y Tversky, A., «On the elicitation of preferences for alternative therapies», *New England Journal of Medicine*, 306, 1982, págs. 1259-1262, <https://doi.org/10.1056/NEJM198205273062103>.

Meehl, P. E., *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*, Brattleboro, Vermont, Echo Point Books, 2013 [1954].

Mellers, B. A., Hertwig, R., y Kahneman, D., «Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration», *Psychological Science*, 12, 2001, págs. 269-275, <https://doi.org/10.1111/1467-9280.00350>.

Mellers, B. A., Ungar, L., Baron, *et al.*, «Psychological strategies for winning a geopolitical forecasting tournament», *Psychological Science*, 25, 2014, págs. 1106-1115, <https://doi.org/10.1177/0956797614524255>.

Mercier, H., *Not Born Yesterday: The Science of Who We Trust and What We Believe*, Princeton (NJ), Princeton University Press, 2020.

Mercier, H., y Sperber, D., «Why do humans reason? Arguments for an argumentative theory», *Behavioral and Brain Sciences*, 34, 2011, págs. 57-111, <https://doi.org/10.1017/S0140525X10000968>.

—, *The Enigma of Reason*, Cambridge (MA), Harvard University Press, 2017.

Mercier, H., Trouche, E., Yama, H., *et al.*, «Experts and laymen grossly underestimate the benefits of argumentation for reasoning», *Thinking & Reasoning*, 21, 2015, págs. 341-355, <https://doi.org/10.1080/13546783.2014.981582>.

Michel, J.-B., Shen, Y. K., Aiden, A. P., *et al.*, «Quantitative analysis of culture using millions of digitalized books», *Science*, 331, 2011, págs. 176-182.

Millenson, J. R., «An inexpensive Geiger gate for controlling probabilities of events», *Journal of the Experimental Analysis of Behavior*, 8, 1965, págs. 345-346.

Miller, J. B., y Sanjurjo, A., «Surprised by the hot hand fallacy? A truth in the law of small numbers», *Econometrica*, 86, 2018, págs. 2019-2047, <https://doi.org/10.3982/ECTA14943>.

—, «A bridge from Monty Hall to the hot hand: The principle of restricted choice», *Journal of Economic Perspectives*, 33, 2019, págs. 144-162, <https://doi.org/10.1257/jep.33.3.144>.

Mischel, W., y Baker, N., «Cognitive appraisals and transformations in delay behavior», *Journal of Personality and Social Psychology*, 31, 1975, págs. 254-261, <https://doi.org/10.1037/h0076272>.

Mlodinow, L., *The Drunkard's Walk: How Randomness Rules Our Lives*, Nueva York, Vintage, 2009 (trad. cast.: *El andar del borracho: cómo el azar gobierna nuestras vidas*, Barcelona, Crítica, 2010).

Moore, D. W., «Three in four Americans believe in paranormal», *Gallup*, 16 de enero de 2005, <https://news.gallup.com/poll/16915/three-four-americans-believe-paranormal.aspx>.

Morewedge, C. K., Yoon, H., Scopelliti, I., *et al.*, «Debiasing decisions: Improved decision making with a single training intervention», *Policy Insights from the Behavioral and Brain Sciences*, 2, 2015, págs. 129-140, <https://doi.org/10.1177/2372732215600886>.

Mueller, J., *Overblown: How Politicians and the Terrorism Industry Inflate National Security Threats, and Why We Believe Them*, Nueva York, Free Press, 2006.

—, *The Stupidity of War: American Foreign Policy and the Case for Complacency*, Nueva York, Cambridge University Press, 2021.

Myers, D. G., *A Friendly Letter to Skeptics and Atheists*, Nueva York, Wiley, 2008.

Nagel, T., *The Possibility of Altruism*, Princeton (NJ), Princeton University Press, 1970 (trad. cast.: *La posibilidad del altruismo*, México, Fondo de Cultura Económica, 2004).

—, *The Last Word*, Nueva York, Oxford University Press, 1997 (trad. cast.: *La última palabra*, Barcelona, Gedisa, 2001).

National Research Council, *The Polygraph and Lie Detection*, Washington, D. C., National Academies Press, 2003.

—, *Strenghtening Forensic Science in the United States: A Path Forward*, Washington, D. C., National Academies Press, 2009.

National Science Board, *Science and Engineering Indicators 2014*, Alexandria (VI) National Science Foundation, 2014, <https://www.nsf.gov/statistics/seind14/index.cfm/home>.

—, *The State of U.S. Science and Engineering 2020*, Alexandria (VI), National Science Foundation, 2020, <https://ncses.nsf.gov/pubs/nsb20201/>.

Nature (grupo editor), «A four-year timeline of Trump's impact on science», *Nature*, 5 de octubre de 2020, <https://doi.org/10.1038/d41586020-02814-3>.

—, «In praise of replication studies and null results», *Nature*, 578, 2020, págs. 489-490, <https://doi.org/10.1038/d41586-020-00530-6>.

Neumann, J. von, y Morgenstern, O., *Theory of Games and Economic Behavior*, Princeton (NJ), Princeton University Press, 2007 [1953].

Nickerson, R. S., «Hempel's paradox and Wason's selection task: Logical and psychological puzzles of confirmation», *Thinking & Reasoning*, 2, 1996, págs. 1-31, <https://doi.org/10.1080/135467896394546>.

—, «Confirmation bias: A ubiquitous phenomenon in many guises», *Review of General Psychology*, 2, 1998, págs. 175-220, <https://doi.org/10.1037/1089-2680.2.2.175>.

Nolan, D., Bremer, M., Tupper, S., *et al.*, *Barnstable County High Crash Locations: Cape Cod Commission*, 2019, <https://www.capecodcommission.org/resourcelibrary/file/?url=/dept/commission/team/tr/Reference/Safety-General/Top50CrashLocs_2018Final.pdf>.

Norberg, J., *Progress: Ten Reasons to Look Forward to the Future*, Londres, Oneworld, 2016 (trad. cast.: *Progreso: diez razones para mirar al futuro con optimismo*, Barcelona, Deusto, 2017).

Nordhaus, W., «Critical assumptions in the Stern Review on climate change», *Science*, 317, 2007, págs. 201-202, <https://doi.org/10.1126/science.1137316>.

Norenzayan, A., Smith, E. E., Kim, B., y Nisbett, R. E., «Cultural preferences for formal versus intuitive reasoning», *Cognitive Science*, 26, 2002, págs. 653-684.

Norman, A., «Why we reason: Intention-alignment and the genesis of human rationality», *Biology and Philosophy*, 31, 2016, págs. 685-704, <https://doi.org/10.1007/s10539-016-9532-4>.

—, *Mental Immunity: Infectious Ideas, Mind Parasites, and the Search for a Better Way to Think*, Nueva York, HarperCollins, 2021.

Nyhan, B., «Building a better correction: Three lessons from new research on how to counter misinformation», *Columbia Journalism Review*, 2013, <http://archives.cjr.org/united_states_project/building_a_better_correction_nyhan_new_misperception_research.php>.

—, «Fake news and bots may be worrisome, but their political power is overblown», *The New York Times*, 13 de febrero de 2018, <https://www.nytimes.com/2008/02/13/upshot/fake-news-and-bots-may-be-worrisome-but-their-political-power-is-overblown.html>.

Nyhan, B., y Reifler, J., *Misinformation and Fact-Checking: Research Findings from Social Science*, Washington, D. C., New America Foundation, 2012.

—, «The roles of information deficits and identity threat in the prevalence of misperceptions», *Journal of Elections, Public Opinion and Parties*, 29, 2019, págs. 222-244, <https://doi.org/10.1080/17457289.2018.1465061>.

O'Keefe, S. M., «One in three Americans would not get COVID-19 vaccine», *Gallup*, 7 de agosto de 2020, <https://gallup.com/poll/317018/one-three-americans-not-covid-vaccine.aspx>.

Open Science Collaboration, «Estimating the reproductibility of psychological science», *Science*, 349, 2015, <https://doi.org/10.1126/science.aac4716>.

Paresky, P., Haidt, J., Strossen, N., y Pinker, S., «*The New York Times* surrendered to an outrage mob. Journalism will suffer for it», *Politico*, 14 de mayo de 2020, <https://www.politico.com/news/magazine/2020/05/14/bret-stephens-new-york-times-outrage-backlash-256494>.

Parker, A. M., Bruine de Bruin, W., Fischhoff, B., y Weller, J., «Robustness of decision-making competence: Evidence from two measures and an 11-year longitudinal study», *Journal of Behavioral Decision Making*, 31, 2018, págs. 380-391, <https://doi.org/10.1002/bdm.2059>.

Pashler, H., y Wagenmakers, E. J., «Editors' introduction to the special section on replicability in psychological science: A crisis of confidence?», *Perspectives on Psychological Science*, 7, 2012, págs. 528-530, <https://doi.org/10.1177/1745691612465253>.

Paulos, J. A., *Innumeracy: Mathematical Illiteracy and Its Consequences*, Nueva York, MacMillan, 1988 (trad. cast.: *El hombre anumérico: el analfabetismo matemático y sus consecuencias*, Barcelona, Tusquets, 2016).

Payne, J. L., *A History of Force: Exploring the Worldwide Movement against Habits of Coercion, Bloodshed, and Mayhem*, Sandpoint (ID), Lytton, 2004.

Pearl, J., *Causality: Models, Reasoning, and Inference*, Nueva York, Cambridge University Press, 2000.

—, y Mackenzie, D., *The Book of Why: The New Science of Cause and Effect*, Nueva York, Basic Books, 2018 (trad. cast.: *El libro del porqué: la nueva ciencia de la causa y el efecto*, Barcelona, Pasado y Presente, 2020).

Pennycook, G., Cannon, T. D., y Rand, D. G., «Prior exposure increases perceived accuracy of fake news», *Journal of Experimental Psychology: General*, 147, 2018, págs. 1865-1880, <https://doi.org/10.1037/xge0000465>.

Pennycook, G., Cheyne, J. A., Koehler, D. J., y Fugelsang, J. A., «On the belief that beliefs should change according to evidence: Implications for conspirational, moral, paranormal, political, religious, and science beliefs», *Judgment and Decision Making*, 15, 2020, págs. 476-498, <https://doi.org/10.31234/osf.io/a7k96>.

Pennycook, G., Cheyne, J. A., Seli, P., *et al*., «Analytic cognitive style predicts religious and paranormal belief», *Cognition*, 123, 2012, págs. 335-346, <https://doi.org/10.1016/j.cognition.2012.03.003>.

Pennycook, G., y Rand, D. G., «The cognitive science of fake news», Psyarxiv, 2020, <https://psyarxiv.com/ar96c>.

—, «Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking», *Journal of Personality*, 88, 2020, <https://doi.org/10.1111/jopy.12476>.

Pew Forum on Religion and Public Life, *Many Americans mix multiple faiths*, Washington, Pew Research Center, 2009, <https://www.pewforum.org/2009/12/09/many-americans-mix-multiple-faiths/>.

Pinker, S., *The Language Instinct*, Nueva York, HarperCollins, 2007 [1994] (trad. cast.: *El instinto del lenguaje: cómo la mente construye el lenguaje*, Madrid, Alianza, 2012).

—, *The Stuff of Thought: Language as a Window into Human Nature*, Nueva York, Viking, 2007 (trad. cast.: *El mundo de las palabras: una introducción a la naturaleza humana*, Barcelona, Paidós, 2007).

—, *How the Mind Works*, Nueva York, W. W. Norton, 2009 [1997] (trad. cast.: *Cómo funciona la mente*, Barcelona, Destino, 2008).

—, «The cognitive niche: Coevolution of intelligence, sociality, and language», *Proceedings of the National Academy of Sciences*, 107, 2010, págs. 8993-8999, <https://doi.org/10.1073/pnas.0914630107>.

—, *Words and Rules: The Ingredients of Language*, Nueva York, HarperCollins, 2011 [1999].

—, *The Better Angels of Our Nature: Why Violence Has Declined*, Nueva York, Viking, 2011 (trad. cast.: *Los ángeles que llevamos dentro: el declive de la violencia y sus implicaciones*, Barcelona, Paidós, 2012).

—, «Why are states so red and blue?», *The New York Times*, 24 de octubre de 2012, <http://opinionator.blogs.nytimes.com/2012/10/24/why-are-states-so-red-and-blue/?_r=0>.

—, «Rock star psychologist Steven Pinker explains why #thedress looked white, not blue», *Forbes*, 28 de febrero de 2015, <https://www.forbes.com/sites/matthewherper/2015/02/28/psychologist-and-author-stephen-pinker-explains-thedress/>.

—, *The Blank Slate: The Modern Denial of Human Nature*, Nueva York, Penguin, 2016 [2002] (trad. cast.: *La tabla rasa: la negación moderna de la naturaleza humana*, Barcelona, Paidós, 2012).

—, *Enlightenment Now: The Case for Reason, Science, Humanism, and Progress*, Nueva York, Viking, 2018 (trad. cast.: *En defensa de la Ilustración: por la razón, la ciencia, el humanismo y el progreso*, Barcelona, Paidós, 2018).

Pinker, S., y Mehler, J. (comps.), *Connections and Symbols*, Cambridge (MA), MIT Press, 1988.

Pinker, S., y Prince, A., «The nature of human concepts: Evidence from an unusual source», en S. Pinker (comp.), *Language, Cognition, and Human Nature: Selected Articles*, Nueva York, Oxford University Press, 2013.

Plato [Platón], «Euthyphro», en J. M. Cooper (comp.), *Plato: Five Dialogues – Euthyphro, Apology, Crito, Meno, Phaedo*, Indianápolis, Hackett, 2002 [399-390 a. C.] (trad. cast.: «Eutifrón», en *Diálogos*, vol. 1, Biblioteca Clásica Gredos, Barcelona, 2019).

Polderman, T. J. C., Benyamin, B., Leeuw, C. A. de, *et al.*, «Meta-analysis of the heritability of human traits base on fifty years of twin studies», *Nature Genetics*, 47, 2015, págs. 702-709, <https://doi.org/10.1038/ng.3285>.

Popper, K. R., *Realism and the Aim of Science*, Londres, Routledge, 1983 (trad. cast.: *Realismo y el objetivo de la ciencia*, Madrid, Tecnos, 2010).

Poundstone, W., *Prisoner's Dilemma: John von Neumann, Game Theory, and the Puzzle of the Bomb*, Nueva York, Anchor, 1992 (trad. cast.: *El dilema del prisionero: John von Neumann, la teoría de juegos y la bomba*, Madrid, Alianza, 2015).

President's Council of Advisors on Science and Technology, *Report to the President: Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*, 2016, <https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf>.

Priest, G., *Logic: A Very Short Introduction*, Nueva York, Oxford University Press, 2017 (2.ª ed.) (trad. cast.: *Una brevísima introducción a la lógica*, México, Océano, 2006).

Proctor, R. N., *The Nazi War on Cancer*, Princeton (NJ), Princeton University Press, 2000.

Pronin, E., Lin, D. Y., y Ross, L., «The bias blind spot: Perceptions of bias in self versus others», *Personality and Social Psychology Bulletin*, 28, 2002, págs. 369-381, <https://doi.org/10.1177/0146167202286008>.

Prooijen, J.-W. van, y Vugt, M. van, «Conspiracy theories: Evolved functions and psychological mechanisms», *Perspectives on Psychological Science*, 13, 2018, págs- 770-788, <https://doi.org/10.1177/1745691618774270>.

Purves, D., y Lotto, R. B., *Why We See What We Do: An Empirical Theory of Vision*, Sunderland (MA), Sinauer, 2003.

Rachels, J., y Rachels, S., *The Elements of Moral Philosophy*, Columbus (OH), McGraw-Hill (6.ª ed.), 2010 (trad. cast.: *Introducción a la filosofía moral*, Madrid, FCE, 2011).

Raemon, «What exactly is the "Rationality Community"?», *LessWrong*, 9 de abril de 2017, <https://www.lesswrong.com/posts/s8yvtCbbZW2S4-WnhE/what-exactly-is-the-rationality-community>.

Railton, P., «Moral realism», *Philosophical Review*, 95, 1986, págs. 163-207, <https://doi.org/10.2307/2185589>.

Rauch, J., «The constitution of knowledge», *National Affairs*, otoño de 2018, <https://www.nationalaffairs.com/publications/detail/the-constitution-of-knowledge>.

—, *The Constitution of Knowledge: A Defense of Truth*, Washington, D. C., Brookings Institution Press, 2021.

Richardson, J., Smith, A., Meaden, S., y Flip Creative, «Thou shalt not commit logical fallacies», 2020, <https://yourlogicalfallacyis.com/>.

Richardson, L. F., *Statistics of Deadly Quarrels*, Pittsburgh, Boxwood Press, 1960.

Ridley, M., *The Origins of Virtue: Human Instincts and the Evolution of Cooperation*, Nueva York, Viking, 1997.

—, *The Rational Optimist: How Prosperity Evolves*, Nueva York, HarperCollins, 2010 (trad. cast.: *El optimista racional: ¿tiene límites la capacidad de progreso de la especie humana?*, Madrid, Taurus, 2011).

Ritchie, H., «Causes of death», *Our World in Data*, 2018, <https://ourworldindata.org/causes-of-death>.

Ritchie, S., *Intelligence: All That Matters*, Londres, Hodders & Stoughton, 2015.

Ropeik, D., *How Risky Is It, Really? Why Our Fears Don't Always Match the Facts*, Nueva York, McGraw-Hill, 2010.

Rosch, E., «Principles of categorization», en E. Rosch y B. B. Lloyd (comps.), *Cognition and Categorization*, Hillsdale (NJ), Erlbaum, 1978.

Rosen, J., «The bloods and the crits», *New Republic*, 9 de diciembre de 1996, <https://newrepublic.com/article/74070/the-bloods-and-the-crits>.

Rosenthal, E. C., *The Complete Idiot's Guide to Game Theory*, Nueva York, Penguin, 2011.

Roser, M., «Economic growth», *Our World in Data*, 2016, <https://ourworldindata.org/economic-growth>.

—, Ortiz-Ospina, E., y Ritchie, H., «Life Expectancy», *Our World in Data*, 2013, <https://ourworldindata.org/life-expectancy>.

—, Ritchie, H., Ortiz-Ospina, E., y Hasell, J., «Coronavirus pandemic (COVID-19)», *Our World in Data*, 2020, <https://ourworldindata.org/coronavirus>.

Rosling, H., *Factfulness: Ten Reasons We're Wrong about the World and Why Things Are Better Than You Think*, Nueva York, Flatiron, 2019 (trad. cast.: *Factfulness: diez razones por las que estamos equivocados sobre el mundo y por qué las cosas están mejor de lo que piensas*, Barcelona, Deusto, 2019).

Roth, G. A., Abate, D., Abate, K. H., *et al.*, «Global, regional, and national agesexspecific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017», *The Lancet*, 392, 2018, págs. 1736-1788, <https://doi.org/10.1016/S0140-6736(18)32203-7>.

Rumelhart, D. E., Hinton, G. E., y Williams, R. J., «Learning representations by back-propagating errors», *Nature*, 323, 1986, págs. 533-536, <https://doi.org/10.1038/323533a0>.

Rumelhart, D. E., McClelland, J. L., y PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1: *Foundations*, Cambridge (MA), MIT Press, 1986.

Rumney, P. N. S., «False allegations of rape», *The Cambridge Law Journal*, 65, 2006, págs. 128-158, <https://doi.org/10.1017/S0008197306007069>.

Russell, B., *Unpopular Essays*, Filadelfia, Routledge, 2009 [1950] (trad. cast.: *Ensayos impopulares*, Barcelona, Edhasa, 2003).

—, «Letter to Mr. Major», en B. Feinberg y R. Kasrils (comps.), *Dear Bertrand Russell: A Selection of his Correspondence with the General Public 1950-1968*, Londres, Allen & Unwin, 1969.

Russett, B., y Oneal, J. R., *Triangulating Peace: Democracy, Interdependence, and International Organizations*, Nueva York, Norton, 2001.

Sá, W.C., West, R. F., y Stanovich, K. E., «The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill», *Journal of Educational Psychology*, 91, 1999, págs. 497-510, <https://doi.org/10.1037/0022-0663.91.3.497>.

Saenen, L., Heyvaert, M., Dooren, W. van, Schaeken, W., y Onghena, P., «Why humans fail in solving the Monty Hall dilemma: A systematic review», *Psychologica Belgica*, 58, 2018, págs. 128-158, <https://doi.org/10.5334/pb.274>.

Sagan, S. D., y Suri, J., «The madman nuclear alert: Secrecy, signaling, and safety in October 1969», *International Security*, 27, 2003, págs. 150-183.

Saldin, R. P., y Teles, S. M., *Never Trump: The Revolt of the Conservative Elites*, Nueva York, Oxford University Press, 2020.

Salganik, M. J., Lundberg, I., Kindel, A. T., *et al.*, «Measuring the predictability of life outcomes with a scientific mass collaboration», *Proceedings of the National Academy of Sciences*, 117, 2020, págs. 8398-8403, <https://doi.org/10.1073/pnas.1915006117>.

Satel, S., *When Altruism Isn't Enough: The Case for Compensating Kidney Donors*, Washington, D. C., AEI Press, 2008.

Savage, I., «Comparing the fatality risks in United States transportation across modes and over time», *Research in Transportation Economics*, 43, 2013, págs. 9-22, <https://doi.org/10.1016/j.retrec.2012.12.011>.

Savage, L. J., *The Foundations of Statistics*, Nueva York, Wiley, 1954.

Savant, M. vos, «Game show problem», *Parade*, 9 de septiembre de 1990, <https://web.archive.org/web/20130121183432/http://marilynvos-savant.com/game-show-problem/>.

Schelling, T. C., *The Strategy of Conflict*, Cambridge (MA), Harvard University Press, 1960 (trad. cast.: *La estrategia del conflicto*, Madrid, Tecnos, 1964).

—, «The intimate contest for self-command», en T. C. Schelling (comp.), *Choice and Consequence: Perspectives of an Errant Economist*, Cambridge (MA), Harvard University Press, 1984.

Schneps, L., y Colmez, C., *Math on Trial: How Numbers Get Used and Abused in the Courtoom*, Nueva York, Basic Books, 2013.

Scott-Phillips, T. C., Dickins, T. E., y West, S. A., «Evolutionary theory and the ultimate-proximate distinction in the human behavioral sciences», *Perspectives on Psychological Science*, 6, 2011, págs. 38-47, <https://doi.org/10.1177/1745691610393528>.

Scribner, S., y Cole, M., «Cognitive consequences of formal and informal education», *Science*, 182, 1973, págs. 553-559, <https://doi.org/10.1126/science.182.4112.553>.

Seebach, L., «The fixation with the last 10 percent of risk», *Baltimore Sun*, 13 de abril de 1994, <https://www.baltimoresun.com/news/bs-xpm-1994-04-13-1994103157-story.html>.

Selvin, S., «A problem in probability», *American Statistician*, 29, 1975, pág. 67, <https://www.jstor.org/stable/2683689>.

Serwer, A., «The greatest money manager of our time», *CCN Money*, 15 de noviembre de 2006, <https://money.cnn.com/magazines/fortune_archive/2006/11/27/8394343/index.htm>.

Shackel, N., «Motte and bailey doctrines», 2014, <https://blog.practicalethics.ox.ac.uk/2014/09/motte-and-bailey-doctrines/>.

Sherman, C., «The shark attack that changed Cape Cod forever», *Boston Magazine*, 14 de mayo de 2019, <https://www.bostonmagazine.com/news/2019/05/14/cape-cod-sharks/>.

Shermer, M., *Why People Believe Weird Things*, Nueva York, Freeman, 1997 (trad. cast.: *Por qué creemos en cosas raras: pseudociencia, confusión y otras supersticiones de nuestro tiempo*, Barcelona, Alba, 2008).

—, «The doping dilemma: Game theory helps to explain the pervasive abuse of drugs in cycling, baseball, and other sports», *Scientific American*, abril de 2008, págs. 32-39.

—, *The Believing Brain: From Ghosts and Gods to Politics and Conspiracies*, Nueva York, St. Martin's Press, 2012.

—, *The Moral Arc: How Science and Reason Lead Humanity Toward Truth, Justice, And Freedom*, Nueva York, Henry Holt, 2015.

—, «COVID-19 conspiracists and their discontents», *Quillette*, 7 de mayo de 2020, <https://quillette.com/2020/05/07/covid-19-conspiracists-and-their-discontents/>.

—, «The top ten weirdest things coutndown», *Skeptic*, 2020, <https://www.skeptic.com/reading_room/the-top-10-weirdest-things/>.

—, «Why people believe conspiracy theories», *Skeptic*, 25, 2020, págs. 12-17.

Shtulman, A., *Scienceblind: Why our Intuitive Theories about the World Are So Often Wrong*, Nueva York, Basic Books, 2017.

Shubik, M., «The dollar auction game: A paradox in noncooperative behavior and escalation», *Journal of Conflict Resolution*, 15, 1971, págs. 109-111, <https://doi.org/10.1177/002200277101500111>.

Simanek, D., «Horse's Teeth», 1999, <https://www.lockhaven.edu/~dsimanek/horse.htm>.

Simmons, J. P., Nelson, L. D., y Simonsohn, U., «False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant», *Psychological Science*, 22, 2011, págs. 1359-1366, <https://doi.org/10.1177/0956797611417632>.

Simon, H. A., «Rational choice and the structure of the environment», *Psychological Review*, 63, 1956, págs. 129-138, <https://doi.org/10.1037/h0042769>.

Singer, P., *The Expanding Circle: Ethics and Sociobiology*, Princeton (NJ), Princeton University Press, 2011 [1981].

Sloman, S. A., «The empirical case for two systems of reasoning», *Psychological Bulletin*, 119, 1996, págs. 3-22, <https://doi.org/10.1037/0033-2909.119.1.3>.

—, y Fernbach, P., *The Knowledge Illusion: Why We Never Think Alone*, Nueva York, Penguin, 2017.

Slovic, P., «Perception of risk», *Science*, 236, 1987, págs. 280-285, <https://doi.org/10.1126/science.3563507>.

—, «"If I look at the mass I will never act": Psychic numbing and genocide», *Judgment and Decision Making*, 2, 2007, págs. 79-95, <https://doi.org/10.1007/978-90-481-8647-1_3>.

Slovic, P., y Tversky, A., «Who accepts Savage's axiom?», *Behavioral Science*, 19, 1974, págs. 368-373, <https://doi.org/10.1002/bs.3830190603>.

Soave, R., «Ezra Klein "completely supports" "terrible" Yes Means Yes law», *Reason*, 13 de octubre de 2014, <https://reason.com/2014/10/13/ezra-klein-completely-supports-terrible/>.

Social Progress Imperative, «2020 Social Progress Index», 2020, <https://www.socialprogress.org/>.

Sowell, T., *A Conflict of Visions: Ideological Origins of Political Struggles*, Nueva York, Quill, 1987.

—, *The Vision of the Anointed: Self-congratulation As a Basis for Social Policy*, Nueva York, Basic Books, 1995.

Sperber, D., «Intuitive and reflective beliefs», *Mind & Language*, 12, 1997, págs. 67-83, <https://doi.org/10.1111/j.1468-0017.1997.tb00062.x>.

—, Cara, F., y Girotto, V., «Relevance theory explains the selection task», *Cognition*, 57, 1995, págs. 31-95, <https://doi.org/10.1016/0010-0277(95)00666-M>.

Spinoza, B., *Ethics*, Nueva York, Oxford University Press, 2000 [1677] (trad. cast.: *Ética demostrada según el orden geométrico*, Madrid, Trotta, 2020).

Stango, V., y Zinman, J., «Exponential growth bias and household finance», *The Journal of Finance*, 64, 2009, págs. 2807-2849, <https://doi.org/10.1111/j.1540-6261.2009.01518.x>.

Stanovich, K. E., «On the distinction between rationality and intelligence: Implications for understanding individual differences in reasoning», en K. J. Holyoak y R. G. Morrison (comps.), *The Oxford Handbook of Thinking and Reasoning*, Nueva York, Oxford University Press, 2012.

—, «How to think rationally about world problems», *Journal of Intelligence*, 6(2), 2018, <https://doi.org/10.3390/jintelligence6020025>.

—, «The bias that divides us», *Quillette*, 26 de septiembre de 2020, <https://quillette.com/2020/09/26/the-bias-that-devides-us/>.

—, *The Bias That Divides Us: The Science and Politics of Myside Thinking*, Cambridge (MA), MIT Press, 2021.

—, y Toplak, M. E., «The need for intellectual diversity in psychological science: Our own studies of actively open-minded thinking as a case study», *Cognition*, 187, págs. 156-166, 2019, <https://doi.org/10.1016/j.cognition.2019.03.006>.

—, West, R. F., y Toplak, M. E., *The Rationality Quotient: Toward a Test of Rational Thinking*, Cambridge (MA), MIT Press, 2016.

—, y West, R. F., «Cognitive ability and variation in selection task performance», *Thinking and Reasoning*, 4, 1998, págs. 193-230.

Statista Research Department, «Beliefs and conspiracy theories in the U.S. – Statistics & Facts», Statista, 13 de agosto de 2019, <https://www.statista.com/topics/5103/beliefs-and-superstition-in-the-us/#dossierSummary_chapter5>.

Stenger, V. J., *Physics and Psychics: the Search for a World Beyond the Senses*, Búfalo, Nueva York, Prometheus, 1990.

Stevenson, B., y Wolfers, J., «Economic growth and subjective well-being: Reassessing the Easterlin Paradox», *Brookings Papers on Economic Activity*, 1-87, 2008, <https://doi.org/10.3386/w14282>.

Stoppard, T., *Jumpers: A Play*, Nueva York, Grove Press, 1972.

Stuart, E. A., «Matching methods for causal inference: A review and a look forward», *Statistical Science*, 25, 2010, págs. 1-21, <https://doi.org/10.1214/09-STS313>.

Suits, B., *The Grasshopper: Games, Life, and Utopia*, Peterborough, Broadview Press, 2014 [1978] (3.ª ed.).

Sunstein, C. R., y Vermeule, A., «Conspiracy theories», *John M. Olin Program in law and Economics Working Papers*, 387, 2008, <https://dx.doi.org/10.2139/ssrn.1084585>.

Swets, J. A., Dawes, R. M., y Monahan, J., «Better decisions through science», *Scientific American*, 283, 2000, págs. 82-87.

Sydnor, J., «(Over)insuring Modest Risks», *American Economic Journal: Applied Economics*, 2, 2010, págs. 177-199.

Sykes, C. J., *How the Right Lost Its Mind*, Nueva York, St. Martin's Press, 2017.

Taber, C. S., y Lodge, M., «Motivated skepticism in the evaluation of political beliefs», *American Journal of Political Science*, 50, 2006, págs. 755-769, <https://doi.org/10.1111/j.1540-5907.2006.00214.x>.

Talwalkar, P., «The taxi-cab problem», *Mind Your Decisions*, 5 de septiembre de 2013, <https://mindyourdecisions.com//blog/2013/09/05/the-taxi-cab-problem/>.

Tate, J., Jenkins, J., Rich, S., *et al*., «Fatal force», *The Washington Post*, consultado el 14 de octubre de 2020, <https://www.washingtonpost.com/graphics/investigations/police-shootings-database/>.

Temple, N., «The possible importance of income and education as covariates in cohort studies that investigate the relationship between diet and disease», *F1000Research*, 4, 2015, pág. 690, <https://doi.org/10.12688/f1000research.6929.2>.

Terry, Q. C., *Golden Rules and Silver Rules of Humanity: Universal Wisdom of Civilization*, Berkeley (CA), AuthorHouse, 2008.

Tetlock, P. E., «Political psychology or politicized psychology: Is the road to scientific hell paved with good moral intentions?», *Political Psychology*, 15, 1994, págs. 509-529, <https://doi.org/10.2307/3791569>.

—, «Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors», *Psychological Review*, 109, 2002, págs. 451-471, <https://doi.org/10.1037/0033-295X.109.3.451>.

—, «Thinking the unthinkable: Sacred values and taboo cognitions», *Trends in Cognitive Sciences*, 7, 2003, págs. 320-324, <https://doi.org/10.1016/S1364-6613(03)00135-9>.

— *Expert Political Judgment: How Good Is It? How Can We Know?*, Princeton (NJ), Princeton University Press, 2009.

—, «All it takes to improve forecasting is keep score», Seminarios sobre Pensamiento a Largo Plazo, San Francisco, 23 de noviembre de 2015.

—, Kristel, O. V., Elson, S. B., Green, M. C., y Lerner, J. S., «The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals», *Journal of Personality and Social Psychology*, 78, 2000, págs. 853-870, <https://doi.org/10.1037/0022-3514.78.5.853>.

—, y Gardner, D., *Superforecasting: The Art and Science of Prediction*, Nueva York, Crown, 2015.

Thaler, R. H., y Sunstein, C. R., *Nudge: Improving Decisions about Health, Wealth, and Happiness*, New Haven, Yale University Press, 2008 (trad. cast.: *Un pequeño empujón [nudge]: el impulso que necesitas para tomar mejores decisiones sobre salud, dinero y felicidad*, Barcelona, Taurus, 2018).

Thomas, K. A., Freitas, J. de, DeScioli, P., y Pinker, S., «Recursive mentalizing and common knowledge in the bystander effect», *Journal of Experimental Psychology: General*, 145, 2016, págs. 621-629, <https://doi.org/10.1037/xge0000153>.

Thomas, K. A., DeScioli, P., Haque, O. S., y Pinker, S., «The psychology of coordination and common knowledge», *Journal of Personality and Social Psychology*, 107, 2014, págs. 657-676, <https://doi.org/10.1037/a0037037>.

Thompson, C., «QAnon is like a game – A most dangerous game», *WIRED Magazine*, 22 de septiembre de 2020, <https://www.wired.com/story/qanon-most-dangerous-multiplatform-game/>.

Thompson, D. A., y Adams, S. L., «The full moon and ED patient volumes: Unearthing a myth», *The American Journal of Emergency Medicine*, 14, 1996, págs. 161-164, <https://doi.org/10.1016/S0735-

6757(96)90124-2>.
Tierney, J., «Behind Monty Hall's doors: Puzzle, debate and answer», *The New York Times*, 21 de julio de 1991, <https://www.nytimes.com/1991/07/21/us/behind-monty-hall-s-doors-puzzle-debate-and-answer.html>.
Tierney, J., y Baumeister, R. F., *The Power of Bad: How the Negativity Effect Rules Us and How We Can Rule It*, Nueva York, Penguin, 2019.
Todd, B., «Introducing longtermism», 2017, <https://80000hours.org/articles/future-generations/>.
Tollefson, J., «How Trump damaged science – And why it could take decades to recover», *Nature*, 586, 5 de octubre de 2020, págs. 190-194, <https://www.nature.com/articles/d41586-020-02800-9>.
Toma, M., «Gen Ed 1066 decision-making competence survey», Universidad de Harvard, 2020.
Tooby, J., Cosmides, L., y Price, M. E., «Cognitive adaptations for *n*-person exchange: The evolutionary roots of organizational behavior», *Managerial and Decision Economics*, 27, 2006, págs. 103-129, <https://doi.org/10.1002/mde.1287>.
Tooby, J., y Cosmides, L., «Ecological rationality and the multimodular mind: Grounding normative theories in adaptive problems», en K. I. Manktelow y D. E. Over (comps.), *Rationality: Psychological and Philosophical Perspectives*, Londres, Routledge, 1993.
Tooby, J., y DeVore, I., «The reconstruction of hominid behavioral evolution through strategic modeling», en W. G. Kinzey (comp.), *The Evolution of Human Behavior: Primate Models*, Albany (NY), Suny Press, 1987.
Toplak, M. E., West, R. F., y Stanovich, K. E., «Real-world correlates of performance on heuristics and biases tasks in a community sample», *Journal of Behavioral Decision Making*, 30, 2017, págs. 541-554, <https://doi.org/10.1002/bdm.1973>.
Trivers, R. L., «The evolution of reciprocal altruism», *The Quarterly Review of Biology*, 46, 1971, págs. 35-57, <https://doi.org/10.1086/406755>.

Tversky, A., «Intransitivity of preferences», *Psychological Review*, 76, 1969, págs. 31-48, <https://doi.org/10.1037/h0026750>.

—, «Elimination by aspects: A theory of choice», *Psychological Review*, 79, 1972, págs. 281-299, <https://doi.org/10.1037/h0032955>.

Tversky, A., y Kahneman, D., «Belief in the law of small numbers», *Psychological Bulletin*, 76, 1971, págs. 105-110, <https://doi.org/10.1037/h0031322>.

—, «Availability: A heuristic for judging frequency and probability», *Cognitive Psychology*, 5, 1973, págs. 207-232, <https://doi.org/10.1016/0010-0285(73)90033-9>.

—, «Judgment under uncertainty: Heuristics and biases», *Science*, 185, 1974, págs. 1124-1131, <https://doi.org/10.1126/science.185-4157.1124>.

—, «The framing of decisions and the psychology of choice», *Science*, 211, 1981, págs. 453-458, <https://doi.org/10.1126/science.7455683>.

—, «Evidential impact of base rates», en D. Kahneman, P. Slovic y A. Tversky (comps.), *Judgment under Uncertainty: Heuristics and Biases*, Nueva York, Cambridge University Press, 1982.

—, «Extensions versus intuitive reasoning: The conjunction fallacy in probability judgment», *Psychological Review*, 90, 1983, págs. 293-315.

Twain, M., *Following the Equator*, Nueva York, Dover, 1989 [1897] (trad. cast.: *Viaje alrededor del mundo, siguiendo el Ecuador*, Barcelona, Laertes, 1992).

Uscinski, J. E., y Parent, J. M., *American Conspiracy Theories*, Nueva York, Oxford University Press, 2014.

Vaci, N., Edelsbrunner, P., Stern, E., *et al.*, «The joint influence of intelligence and practice on skill development throughout the life span», *Proceedings of the National Academy of Sciences*, 116, 2019, págs. 18363-18369, <https://doi.org/10.1073/pnas.1819086116>.

VanderWeele, T. J., «Commentary: Resolutions of the birthweight paradox: competing explanations and analytical insights», *International Journal of Epidemiology*, 43, 2014, págs. 1368-1373, <https://doi.org/10.1093/ije/dyu162>.

Varian, H., «Recalculating the costs of global climate change», *The New York Times*, 14 de diciembre de 2006, <https://nytimes.com/2006/12/14/business/14scene.html>.

Vazsonyi, A., «Which door has the Cadillac?», *Decision Line*, 1999, págs. 17-19, <https://web.archive.org/web/20140413131827/http://www.decisionsciences.org/DecisionLine/Vol30/30_1/vazs30_1.pdf>.

Venkataraman, B., *The Optimist's Telescope: Thinking Ahead in a Reckless Age*, Nueva York, Riverhead Books, 2019.

Vosoughi, S., Roy, D., y Aral, S., «The spread of true and false news online», *Science*, 359, 2018, págs. 1146-1151, <https://doi.org/10.1126/science.aap9559>.

Wagenaar, W. A., y Sagaria, S. D., «Misperception of exponential growth», *Perception & Psychophysics*, 18, 1975, págs. 416-422, <https://doi.org/10.3758/BF03204114>.

Wagenaar, W. A., y Timmers, H., «The pond-and-duckweed problem: Three experiments on the misperception of exponential growth», *Acta Psychologica*, 43, 1979, págs. 239-251, <https://doi.org/10.1016/0001-6918(79)90028-3>.

Walker, C., Petulla, S., Fowler, K., *et al*., «10 years. 180 school shootings. 356 victims», CNN, julio de 2019, <https://www.cnn.com/interactive/2019/07/us/ten-years-of-school-shootings-trnd/>.

Wan, W., y Shammas, B., «Why Americans are numb to be the staggering coronavirus death toll», *The Washington Post*, 21 de diciembre de 2020, <https://www.washingtonpost.com/health/2020/12/21/covid-why-we-ignore-deaths/>.

Warburton, N., *Thinking from A to Z*, Nueva York, Routledge, 2007 (3.ª ed.) (trad. cast.: *Pensar de la A a la Z*, Barcelona, Gedisa, 2016).

Wason, P. C., «Reasoning», en B. M. Foss (comp.), *New Horizons in Psychology*, Londres, Penguin, 1966.

Weber, M., *Economy and Society: A New Translation*, Cambridge (MA), Harvard University Press, 2019 [1922] (trad. cast.: *Economía y sociedad: esbozo de sociología comprensiva*, Madrid, Fondo de

Cultura Económica, 2002).
Weissman, M. B., «Do GRE scores help predict getting a physics Ph.D.? A comment on a paper by Miller et al.», *Science Advances*, 6, 2020, eaax3787, <https://doi.org/10.1126/sciadv.aax3787>.
Welzel, C., *Freedom Rising: Human Empowerment and the Quest for Emancipation*, Nueva York, Cambridge University Press, 2013.
Wilkinson, W., *The Density Divide: Urbanization, Polarization, and Populist Backlash*, Washington, D. C., Niskanen Center, 2019, <https://www.niskanencenter.org/the-density-divide-urbanization-polarization-and-populist-backlash/>.
Williams, D., «Motivated ignorance, rationality, and democratic politics», *Synthese*, 2020, págs. 1-21.
Willingham, D. T., «Critical thinking: Why is it so hard to teach?», *American Educator*, 31, 2007, págs. 8-19, <https://doi.org/10.3200/AEPR.109.4.21-32>.
Wittgenstein, L., *Philosophical Investigations*, Nueva York, Macmillan, 1953 (trad. cast.: *Investigaciones filosóficas*, Madrid, Trotta, 2017).
Wolfe, D., y Dale, D., «“It’s going to disappear”: A timeline of Trump’s claims that Covid-19 will vanish», CNN, 31 de octubre de 2020, <https://www.cnn.com/interactive/2020/10/politics/covid-disappearing-trump-comment-tracker/>.
Wolfe, J. M., Kluender, K. R., Levi, D. M., *et al.*, *Sensation & Perception*, Sunderland (MA), Sinauer, 2020 (6.ª ed.).
Wollstonecraft, M., *A Vindication of the Rights of Woman with Structures on Political and Moral Subjects*, Nueva York, Cambridge University Press, 1995 [1792] (trad. cast.: *Vindicación de los derechos de la mujer*, Barcelona, Penguin Clásicos, 2020).
Wood, T., y Porter, E., «The elusive backfire effect: Mass attitudes’ steadfast factual adherence», *Political Behavior*, 41, 2019, págs. 135-163, <https://doi.org/10.1007/s11109-018-9443-y>.
Yang, A., sitio web oficial de la campaña de Yang, 2020, <www.yang2020.com>.

Yglesias, M., «Defund police is a bad idea, not a bad slogan», *Slow Boring*, 7 de diciembre de 2020, <https://www.slowboring.com/p/defund-police-is-a-bad-idea-not-a>.

—, «The End of Policing left me convinced we still need policing», *Vox*, 18 de junio de 2020, <https://www.vox.com/2020/6/18/21293784/alex-vitale-end-of-policing-review>.

Young, C., «The argument against affirmative consent laws gets Voxjacked», *Reason*, 15 de octubre de 2014, <https://reason.com/2014/10/15/the-argumentagainstaffirmative-consent/>.

—, «Crying rape», *Slate*, 18 de septiembre de 2014, <https://slate.com/human-interest/2014/09/false-rape-accusations-why-must-we-pretend-they-never-happen.html>.

Zelizer, V. A., *The Purchase of Intimacy*, Princeton (NJ), Princeton University Press, 2005 (trad. cast.: *La negociación de la intimidad*, México, Fondo de Cultura Económica, 2010).

Ziman, J. M., *Reliable Knowledge: An Exploration of the Grounds for Belief in Science*, Nueva York, Cambridge University Press, 1978.

Zutter, A. de, Horselenberg, R., y Koppen, P. J. van, «The prevalence of false allegations of rape in the United States from 2006-2010», *Journal of Forensic Psychology*, 2, 2017, <https://doi.org/10.4172/2475-319X.1000119>.

Notas

1. Russell, *Unpopular Essays*, Filadelfia, Routledge, 2009 [1950] (trad. cast.: *Ensayos impopulares*, Barcelona, Edhasa, 2003).

2. Spinoza, *Ethics*, Nueva York, Oxford University Press, 2000 [1677] (trad. cast.: *Ética demostrada según el orden geométrico*, Madrid, Trotta, 2020), «Prefacio».

3. Datos sobre el progreso humano: Pinker, *Enlightenment Now: The Case for Reason, Science, Humanism, and Progress*, Nueva York, Viking, 2018 (trad. cast.: *En defensa de la Ilustración: por la razón, la ciencia, el humanismo y el progreso*, Barcelona, Paidós, 2018).

4. Sobre los sans del Kalahari: Lee, B., y Daly, R. (comps.), *The Cambridge Encyclopedia of Hunters and Gatherers*, Cambridge (RU), Cambridge University Press, 1999. Los sans, previamente conocidos como bosquimanos, comprenden los pueblos ju/'hoan (anteriormente !kung), tuu, gana, gwi y khoi, que se escriben de varias maneras.

5. Marlowe, *The Hadza: Hunter-gatherers of Tanzania*, Berkeley, University of California Press, 2010.

6. Liebenberg trabaja con los sans !xõ/gwi, khomani y ju/’hoan (anteriormente !kung). Los ejemplos incluidos aquí son de los !xõ. Las experiencias de Liebenberg con los sans, así como su teoría de que el pensamiento científico evolucionó a partir del rastreo, se exponen en Liebenberg, *The Origin of Science: The Evolutionary Roots of Scientific Reasoning and its Implications for Tracking Science*, Cape Town, CyberTracker, 2021 [2013], <https://cybertracker.org/downloads/tracking/Liebenberg-2013-The-Origin-of-Science.pdf>; *The Art of Tracking: The Origin of Science*, Cape Town, David Philip, 1990; y Liebenberg, Am//Ao y Lombard, M., «Tracking Science: An Alternative for Those Excluded by Citizen Science», *Citizen Science Theory and Practice*, 6(1), 2021. Los ejemplos adicionales son de Liebenberg, «Notes on tracking and trapping: Examples of hunter-gatherer ingenuity», manuscrito inédito, 2020, <https://stevenpinker.com/files/pinker/files/liebenberg.pdf>. Para otras descripciones de la racionalidad de los cazadores recolectores, véase Chagnon, *Yanomamö*, Fort Worth, Texas, Harcourt Brace, 1997 (5.ª ed.); Kingdon, *Self-Made Man: Human Evolution from Eden to Extinction?*, Nueva York, Wiley, 1993; Marlowe, *The Hadza, op. cit*.

7. Un vídeo de una caza por persistencia, narrado por David Attenborough, puede verse aquí: <https://youtu.be/826HMLoiE_o>.

8. Liebenberg, *The Origin of Science*, *op. cit*., pág. 57.

9. Comunicación personal de Louis Liebenberg, 11 de agosto de 2020.

10. Liebenberg, *The Origin of Science*, *op. cit*., pág. 104.

11. Liebenberg, «Notes on tracking.», art. cit., y comunicación personal, 27 de mayo de 2020.

12. Moore, «Three in four Americans believe in paranormal», *Gallup*, 16 de enero de 2005, <https://news.gallup.com/poll/16915/three-fouramericans-believe-paranormal.aspx>. Véase también Pew Forum on Religion and Public Life, *Many Americans mix multiple faiths*, Washington, Pew Research Center, 2009, <https://www.pewforum.org/2009/12/09/many-americans-mix-multiple-faiths/>, y nota 8 del capítulo 10, más abajo.

13. Vosoughi, Roy y Aral, «The spread of true and false news online», *Science*, 359, 2018, págs. 1146-1151, <https://doi.org/10.1126/science.aap9559>.

14. Pinker, «The cognitive niche: Coevolution of intelligence, sociality, and language», *Proceedings of the National Academy of Sciences*, 107, 2010, págs. 8993-8999, <https://doi.org/10.1073/pnas.0914630107>; Tooby y DeVore, «The reconstruction of hominid behavioral evolution through strategic modeling», en W. G. Kinzey (comp.), *The Evolution of Human Behavior: Primate Models*, Albany (NY), Suny Press, 1987.

15. Amos Tversky (1937-1996) y Daniel Kahneman (1934) fueron pioneros en el estudio de las ilusiones y los sesgos cognitivos; véase Tversky y Kahneman, «Judgment under uncertainty: Heuristics and biases», *Science*, 185, 1974, págs. 1124-1131, <https://doi.org/10.1126/science.185-4157.1124>; Kahneman, Solovic y Tversky, *Judgment under Uncertainty: Heuristics and Biases*, Nueva York, Cambridge University Press, 1982; Hastie y Dawes, *Rational Choice in an Uncertain World: The Psychology of Judgment and Decision Making*, Los Ángeles, Sage, 2010 (2.ª ed.), y el superventas de Kahneman *Thinking Fast and Slow* (*Pensar rápido, pensar despacio*, Barcelona, Debolsillo, 2011). Sus vidas y su colaboración se describen en Michael Lewis, *The Undoing Project: A Friendship That Changed Our Minds*, Nueva York, Norton, 2016 (trad. cast.: *Deshaciendo errores: Kahneman, Tversky y la amistad que nos enseñó cómo funciona la mente*, Barcelona, Debate, 2017), y en la declaración autobiográfica de Kahneman con motivo de su Premio Nobel de 2002 (Kahneman, «Daniel Kahneman-facts», The Nobel Prize, 2002, <https://www.nobelprize.org/prizes/economic-sciences/2002/kahneman/facts/>).

16. Frederick, «Cognitive reflection and decision making», *Journal of Economic Perspectives*, 19, 2005, págs. 25-42, <https://doi.org/10.1257/089533005775196732>.

17. Los psicólogos Philip Maymin y Ellen Langer han demostrado que con solo pedir a los individuos que fuesen conscientes de su entorno visual se reducían los errores de razonamiento en diecinueve de los veintidós problemas clásicos de la literatura de la psicología cognitiva.

18. Frederick, «Cognitive reflection...», art. cit.

19. *Ibidem*, pág. 28. En realidad, «un plátano y un panecillo cuestan treinta y siete centavos. El plátano cuesta trece centavos más que el panecillo. ¿Cuánto cuesta el panecillo?».

20. Wagenaar y Sagaria, «Misperception of exponential growth», *Perception & Psychophysics*, 18, 1975, págs. 416-422, <https://doi.org/10.3758/BF03204114>; Wagenaar y Timmers, «The pond-and-duckweed problem: Three experiments on the misperception of exponential growth», *Acta Psychologica*, 43, 1979, págs. 239-251, <https://doi.org/10.1016/00016918(79)90028-3>.

21. Goda, Levy, Manchester *et al.*, «The role of time preferences and exponential-growth bias in retirement savings», *National Bureau of Economic Research Working Paper Series*, 21482, 2015, <https://doi.org/10.3386//w21482>; Stango y Zinman, «Exponential growth bias and household finance», *The Journal of Finance*, 64, 2009, págs. 2807-2849, <https://doi.org/10.1111/j.1540-6261.2009.01518.x>.

22. Citas omitidas para ahorrar la vergüenza a dos amigos.

23. Muertes en Estados Unidos (media móvil de siete días): Roser, Ritchie, Ortiz-Ospina y Hasell, «Coronavirus pandemic (COVID-19)», *Our World in Data*, 2020, <https://ourworldindata.org/coronavirus>, consultado el 23 de agosto de 2020. Causas de mortalidad en Estados Unidos: Ritchie, «Causes of death», *Our World in Data*, 2018, <https://ourworldindata.org/causes-of-death>, consultado el 23 de agosto de 2020; los datos son de 2017.

24. Wason, «Reasoning», en B. M. Foss (comp.), *New Horizons in Psychology*, Londres, Penguin, 1966; véase también Cosmides, «The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task», *Cognition*, 31, 1989, págs. 187-276, <https://doi.org/10.1016/0010-0277(89)90023-1>; Fiddick, Cosmides y Tooby, «No interpretation without representation: The role of domainspecific representations and inferences in the Wason selection task», *Cognition*, 77, 2000, págs. 1-79, <https://doi.org/10.1016/S0010-0277(00)00085-8>; Mercier y Sperber, «Why do humans reason? Arguments for an argumentative theory», *Behavioral and Brain Sciences*, 34, 2011, págs. 57-111, <https://doi.org/10.1017/S0140525X10000968>; Nickerson, «Hempel's paradox and Wason's selection task: Logical and psychological puzzles of confirmation», *Thinking & Reasoning*, 2, 1996, págs. 1-31, <https://doi.org/10.1080/135467896394546>; Sperber, Cara y Girotto, «Relevance theory explains the selection task», *Cognition*, 57, 1995, págs. 31-95, <https://doi.org/10.1016/0010-0277(95)00666M>.

25. Van Benthem, «Logic and reasoning: Do the facts matter?», *Studia Logica*, 88, 2008, págs. 67-84, <https://doi.org/10.1007/s11225-008-91011>, pág. 11.

26. Dado que, en términos lógicos, la elección de P podría refutar la regla con la misma facilidad que la elección de «NO Q», la explicación en términos del sesgo de confirmación es algo más sutil: los participantes despliegan razonamientos para justificar su elección intuitiva inicial, cualquiera que esta sea; véase Nickerson, «Confirmation bias: A ubiquitous phe nomenon in many guises», *Review of General Psychology*, 2, 1998, págs. 175-220, <https://doi.org/10.1037/1089-2680.2.2.175>, y Mercier y Sperber, «Why do humans reason?...», art. cit. Sobre vencer en las discusiones: Dawson, Gilovich y Regan, «Motivated reasoning and performance on the Wason selection task», *Personality and Social Psychology Bulletin*, 28, 2002, págs. 1379-1387, <https://doi.org/10.1177/014616702236869>; Mercier y Sperber, *The Enigma of Reason*, Cambridge (MA), Harvard University Press, 2017.

27. Citado en Grayling, *Toward the Light of Liberty: The Struggles for Freedom and Rights that Made the Modern Western World*, Nueva York, Walker, 2007, pág. 102.

28. Bacon, *Novum organum*, Seattle, Washington, CreateSpace, 2017 [1620] (trad. cast.: *La gran restauración [Novum organum]*, Madrid, Tecnos, 2011).

29. Popper, *Realism and the Aim of Science*, Londres, Routledge, 1983 (trad. cast.: *Realismo y el objetivo de la ciencia*, Madrid, Tecnos, 2010). La tarea de Wason versus la comprobación de hipótesis científicas: Nickerson, «Hempel’s paradox...», art. cit.

30. Sobre la peculiaridad de la tarea de selección: Nickerson, «Hempel's paradox...», art. cit.; Sperber, Cara y Girotto, «Relevance theory...», art. cit.

31. Cheng y Holyoak, «Pragmatic reasoning schemas», *Cognitive Psychology*, 17, 1985, págs. 391-416, <https://doi.org/10.1016/0010-0285(85)900143>; Cosmides, «The logic of social exchange...», art. cit.; Fiddick, Cosmides y Tooby, «No interpretation without representation...», art. cit.; Stanovich y West, «Cognitive ability and variation in selection task performance», *Thinking and Reasoning*, 4, 1998, págs. 193-230. Una perspectiva diferente: Sperber, Cara y Girotto, «Relevance theory...», art. cit.

32. Sobre la racionalidad ecológica, Gigerenzer, «Ecological intelligence: An adaptation for frequencies», en D. D. Cummins y C. Allen (comps.), *The Evolution of Mind*, Nueva York, Oxford University Press, 1998; Tooby y Cosmides, «Ecological rationality and the multimodular mind: Grounding normative theories in adaptive problems», en K. I. Manktelow y D. E. Over (comps.), *Rationality: Psychological and Philosophical Perspectives*, Londres, Routledge, 1993; véase Pinker, *How the Mind Works*, Nueva York, W. W. Norton, 2009 [1997] (trad. cast.: *Cómo funciona la mente*, Barcelona, Destino, 2008), págs. 302-306.

33. El problema fue planteado por el matemático recreativo Martin Gardner («Problems involving questions of probability and ambiguity», *Scientific American*, 201, 1959, págs. 174-182), quien lo denominó el problema de los tres prisioneros; el nombre de Monty Hall se lo puso el estadístico Steven Selvin, «A problem in probability», *American Statistician*, 29, 1975, pág. 67, <https://www.jstor.org/stable/2683689>.

34. Granberg y Brown, «The Monty Hall dilemma», *Personality & Social Psychology Bulletin*, 21, 1995, págs. 711-723, <https://doi.org/10.1177/0146167295217006>; Saenen, Heyvaert, Dooren *et al*., «Why humans fail in solving the Monty Hall dilemma: A systematic review», *Psychologica Belgica*, 58, 2018, págs. 128-158, <https://doi.org/10.5334/pb.274>.

35. Crockett, «The time everyone “corrected” the world’s smartest woman», *Priceonomics*, 19 de febrero de 2015, <https://priceonomics.com/the-time-everyone-corrected-the-worlds-smartest/>; Granberg y Brown, «The Monty Hall dilemma», art. cit.; Tierney, «Behind Monty Hall’s doors: Puzzle, debate and answer», *The New York Times*, 21 de julio de 1991, <https://www.nytimes.com/1991/07/21/us/behind-monty-hall-s-doorspuzzle-debate-and-answer.html>; Savant, «Game show problem», *Parade*, 9 de septiembre de 1990, <https://web.archive.org/web/20130121183432/http://marilynvossavant.com/game-show-problem/>.

36. Crockett, «The time everyone “corrected”...», art. cit.

37. Vazsonyi, «Which door has the Cadillac?», *Decision Line*, 1999, págs. 17-19, <https://web.archive.org/web/20140413131827/http://www.decisionsciences.org/DecisionLine/Vol30/30_1/vazs30_1.pdf>. Mi número de Erdös es 3, gracias a Michel, Shen, Aiden *et al*., «Quantitative analysis of culture using millions of digitalized books», *Science*, 331, 2011, págs. 176-182. El informático Peter Norvig ha redactado un informe con su colega informática (y coautora de Erdös) Maria Klawe.

38. En honor a la verdad, los análisis normativos del dilema de Monty Hall han inspirado voluminosos comentarios y discrepancias; véase <https://en.wikipedia.org/wiki/Monty_Hall_problem>.

39. Pruébalo: Math Warehouse, «Monty Hall Simulation Online», <https://www.mathwarehouse.com/monty-hall-simulation-online/>.

40. Tales como *Late Night with David Letterman*, <https://www.youtube.com/watch?v=EsGc3jC9yas>.

41. Vazsonyi, «Which door...?», art. cit.

42. Sugerido por Granberg y Brown, «The Monty Hall dilemma», art. cit.

43. Reglas de la conversación: Grice, «Logic and conversation», en P. Cole y J. L. Morgan (comps.), *Syntax & Semantics*, vol. 3, *Speech Acts*, Nueva York, Academic Press, 1975; Pinker, *The Stuff of Thought: Language as a Window into Human Nature*, Nueva York, Viking, 2007 (trad. cast.: *El mundo de las palabras: una introducción a la naturaleza humana*, Barcelona, Paidós, 2007), cap. 8.

44. Sobre la historia y los conceptos de *probabilidad*, Gigerenzer, Swijtink, Porter *et al.*, *The Empire of Chance: How Probability Changed Science and Everyday Life*, Nueva York, Cambridge University Press, 1989.

45. Savant, «Game show problem», art. cit.

46. Gracias a Julian de Freitas por llevar a cabo y analizar el estudio. El diseño era similiar al resumido informalmente en Tversky y Kahneman, «Extensions versus intuitive reasoning: The conjunction fallacy in probability judgment», *Psychological Review*, 90, 1983, págs. 307-308. Los ítems se eligieron aquí a partir de un conjunto más grande probado previamente en un estudio piloto. Las diferencias se hallaron en las comparaciones de las calificaciones que dieron los participantes a la conjunción o a uno de sus componentes antes de haber visto el otro (esto es, en una comparación entre los participantes). Cuando comparamos las calificaciones de ambos ítems por los mismos participantes (una comparación para cada participante), la falacia de la conjunción se veía únicamente con los ítems de Rusia y Venezuela. No obstante, el 86 % de los participantes cometían al menos un error de conjunción y, con todos los ítems, una mayoría de los participantes calificaban la probabilidad de la conjunción como igual o mayor que la probabilidad de cada componente.

47. Donaldson, Doubleday, Hefferman, *et al*., «Are talking heads blowing hot air? An analysis of the accuracy of forecasts in the political arena», Hamilton College, 2011, <https://www.hamilton.edu/documents/Analysis-of-Forecast-Accuracy-in-the-Political-Media.pdf>; Tetlock y Gardner, *Superforecasting: The Art and Science of Prediction*, Nueva York, Crown, 2015.

48. Kaplan, «The Coming Anarchy», *The Atlantic*, 1994, <https://www.theatlantic.com/magazine/archive/1994/02/the-coming-anarchy/304670/>.

49. Descensos en guerras, delitos, pobreza y enfermedades: Pinker, *The Better Angels of Our Nature: Why Violence Has Declined*, Nueva York, Viking, 2011 (trad. cast.: *Los ángeles que llevamos dentro: el declive de la violencia y sus implicaciones*, Barcelona, Paidós, 2012); y *Enlightenment Now, op. cit*.

50. Tversky y Kahneman, «Extensions versus intuitive reasoning...», art. cit.

51. Gould, «The streak of streaks», *The New York Review of Books*, 1988, <https://www.nybooks.com/articles/1988/08/18/the-streak-of-streaks/>.

52. Citado por Tversky y Kahneman, «Extensions versus intuitive reasoning...», art. cit., pág. 308.

53. *Ibidem*, pág. 313.

54. Citado en Hertwig y Gigerenzer, «The “conjuction fallacy” revisited: How intelligent inferences look like reasoning errors», *Journal of Behavioral Decision Making*, 12, 1999, págs. 275-305, <https://doi.org/10.1002/(SICI)1099-0771(199912)12:4<275::AID-BDM323>3.0.CO;2-M>.

55. *Ibidem*.

56. *Ibidem*; Tversky y Kahneman, «Extensions versus intuitive reasoning...», art. cit.

57. Kahneman y Tversky, «On the reality of cognitive illusions», *Psychological Review*, 103, 1996, págs. 582-591, <https://doi.org/10.1037/0033295X.103.3.582>.

58. Mellers, Hertwig y Kahneman, «Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration», *Psychological Science*, 12, 2001, págs. 269-275, <https://doi.org/10.1111/14679280.00350>.

59. Purves y Lotto, *Why We See What We Do: An Empirical Theory of Vision*, Sunderland (MA), Sinauer, 2003.

60. Sobre los fracasos de la inteligencia artificial, Marcus y Davis, *Rebooting AI: Building Artificial Intelligence We Can Trust*, Nueva York, Penguin Random House, 2019.

61. Pinker, *How the Mind Works*, *op. cit*., caps. 1 y 4.

62. Pinker, «Rock star psychologist Steven Pinker explains why #thedress looked white, not blue», *Forbes*, 28 de febrero de 2015, <https://www.forbes.com/sites/matthewherper/2015/02/28/psychologist-andauthor-stephen-pinker-explains-thedress/>.

63. Federal Aviation Administration, *Pilot’s Handbook of Aeronautical Knowledge*, Oklahoma City, U. S. Department of Transportation, 2016, <https://www.faa.gov/regulations_policies/handbooks_manuals/aviation/phak/media/pilot_handbook.pdf>, cap. 17.

* Niño mascota ficticio de la portada de la revista estadounidense de humor *MAD [N. del T.]*.

* *Misunderestimate* es uno de los términos incorrectos empleado por G. W. Bush. En este caso, al verbo *underestimat*e («subestimar») se le antepone incorrectamente el prefijo *mis*, común en verbos como *misunderstand* («malinterpretar») o *mislead* («engañar») *[N. del T.]*.

1. Sobre la creencia verdadera justificada y contraejemplos de que es necesaria, pero no suficiente, para el conocimiento: Gettier, «Is justified true belief knowledge?», *Analysis*, 23, 1963, págs. 121-123; Ichikawa y Steup, «The analysis of knowledge», en E. N. Zalta (comp.), *The Stanford Encyclopedia of Philosophy*, 2018, <https://plato.stanford.edu/entries/knowledge-analysis/>

2. James, *The Principles of Psychology*, Nueva York, Dover, 1950 [1890] (trad. cast.: *Principios de psicología*, México, FCE, 1989).

3. Carroll, «What the tortoise said to Achilles», *Mind*, 4, 1895, págs. 178-180.

4. Simplemente se hace: Fodor, *Psychological Explanation: An Introduction to the Philosophy of Psychology*, Nueva York, Random House, 1968 (trad. cast: *La explicación psicológica*, Madrid, Cátedra, 1991); Pinker, *How the Mind Works, op. cit*., cap. 2.

5. Nagel, *The Last Word*, Nueva York, Oxford University Press, 1997 (trad. cast.: *La última palabra*, Barcelona, Gedisa, 2001).

6. Myers, *A Friendly Letter to Skeptics and Atheists*, Nueva York, Wiley, 2008.

7. Pueden verse numerosos ejemplos en las fuentes de la nota 79, cap. 10.

8. Stoppard, *Jumpers: A Play*, Nueva York, Grove Press, 1972, pág. 30.

9. Hume, *A Treatise of Human Nature*, Nueva York, Oxford University Press, 2000 [1739] (trad. cast.: *Tratado de la naturaleza humana: autobiografía*, Madrid, Tecnos, 2002), libro II, parte III, «On the influencing motives of the will».

10. Cohon, «Hume's moral philosophy», en E. N. Zalta (comp.), *The Stanford Encyclopedia of Philosophy*, 2018, <https://plato.stanford.edu/entries/hume-moral/>.

11. Aunque eso no es lo que creía literalmente sobre el gusto en el arte y en el vino, tal como se expresa en «Of the standards of taste» (Gracyk, «Hume's aesthetics», en E. N. Zalta (comp.), *Stanford Encyclopedia of Philosophy*, 2020, <https://plato.stanford.edu/archives/sum2020/entries/humeaesthetics/>). Su tesis a este respecto era únicamente que las metas son intrínsecamente subjetivas.

12. Bob Dylan, «Mr. Tambourine Man».

13. Pinker, *How the Mind Works*, *op. cit.*; Scott-Phillips, Dickins y West, «Evolutionary theory and the ultimate-proximate distinction in the human behavioral sciences», *Perspectives on Psychological Science*, 6, 2011, págs. 38-47, <https://doi.org/10.1177/1745691610393528>.

14. Ainslie, *Breakdown of Will*, Nueva York, Cambridge University Press, 2001; Schelling, «The intimate contest for self-command», en T. C. Schelling (comp.), *Choice and Consequence: Perspectives of an Errant Economist*, Cambridge (MA), Harvard University Press, 1984.

15. Mischel y Baker, «Cognitive appraisals and transformations in delay behavior», *Journal of Personality and Social Psychology*, 31, 1975, págs. 254261, <https://doi.org/10.1037/h0076272>.

16. Ainslie, *Breakdown of Will*, *op. cit.*; Laibson, «Golden eggs and hyperbolic discounting», *Quarterly Journal of Economics*, 112, 1997, págs. 443-477, <https://doi.org/10.1162/003355397555253>; Schelling, «The intimate contest for self-command», art. cit. Véase también Pinker, *The Better Angels...*, *op. cit.*, cap. 9, «Self-control».

17. Frederick, «Cognitive reflection...», art. cit.

18. Jeszeck, Collins, Glickman, Hoffrey y Grover, «Retirement security: Most households approaching retirement have low savings», *United States Government Accountability Office*, 2015, <https://www.gao.gov/assets/680/670153.pdf>.

19. Dasgupta, «The Stern Review’s economics of climate change», *National Institute Economic Review*, 199, 2007, págs. 4-7, <https://doi.org/10.1177/0027950107077111>; Nordhaus, «Critical assumptions in the Stern Review on climate change», *Science*, 317, 2007, págs. 201-202, <https://doi.org/10.1126/science.1137316>; Varian, «Recalculating the costs of global climate change», *The New York Times*, 14 de diciembre de 2006, <https://nytimes.com/2006/12/14/business/14scene.html>; Venkataraman, *The Optimist’s Telescope: Thinking Ahead in a Reckless Age*, Nueva York, Riverhead Books, 2019.

20. MacAskill, *Doing Good Better: Effective Altruism and How You Can Make a Difference*, Nueva York, Penguin, 2015; Todd, «Introducing longtermism», 2017, <https://80000hours.org/articles/future-generations/>.

21. Venkataraman, *The Optimist’s Telescope...*, *op. cit*.

22. Ainslie, *Breakdown of Will, op. cit*; Laibson, «Golden eggs...», art. cit.

23. McClure, Laibson, Loewenstein y Cohen, «Separate neural systems value immediate and delayed monetary rewards», *Science*, 306, 2004, págs. 503-507, <https://doi.org/10.1126/science.1100907>.

24. Homero, *The Odyssey*, Nueva York, Norton, 2018 [700 a. C.] (trad. cast.: *Odisea*, Barcelona, Nueva Biblioteca Clásica Gredos, 2019).

25. Baumeister y Tierney, *Willpower: Rediscovering the Greatest Human Strength*, Londres, Penguin, 2012.

26. Sobre los *nudges* (pequeños empujones) y las «ideas conductuales», Hallsworth y Kirkman, *Behavioral Insights*, Cambridge (MA), MIT Press, 2020; Thaler y Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness*, New Haven, Yale University Press, 2008 (trad. cast.: *Un pequeño empujón [nudge]: el impulso que necesitas para tomar mejores decisiones sobre salud, dinero y felicidad*, Barcelona, Taurus, 2018). Escépticos respecto de los *nudges*: Gigerenzer, «On the supposed evidence for libertarian paternalism», *Review of Philosophy and Psychology*, 6, 2015, págs. 361-383, <https://doi.org/10.1007/s13164-015-0248-1>; Kahan, «Ideology, motivated reasoning, and cognitive reflection», *Judgment and Decision Making*, 8, 2013, págs. 407-424, <http://dx.doi.org/10.2139/ssrn.2182588>.

27. Sobre la ignorancia racional: Gigerenzer, «Gigerenzer's Law of indispensable ignorance», *Edge*, 2004, <https://www.edge.org/response-detail/10224>; Gigerenzer y García-Retamero, «Cassandra's regret: The psychology of not wanting to know», *Psychological Review*, 124, 2017, págs. 179-196; Hertwig y Engel, «Homo ignorans: Deliberately choosing not to know», *Perspectives on Psychological Science*, 11, 2016, págs. 359-372; Williams, «Motivated ignorance, rationality, and democratic politics», *Synthese*, 2020, págs. 1-21; véase también Pinker, *The Stuff of Thought...*, *op. cit*., págs. 422-425.

28. Schelling, *The Strategy of Conflict*, Cambridge (MA), Harvard University Press, 1960 (trad. cast.: *La estrategia del conflicto*, Madrid, Tecnos, 1964).

29. Sobre el juego del gallina, J. S. Goldstein, «Chicken dilemmas: crossing the road to cooperation», en I. W. Zartman y S. Touval (comps.), *International Cooperation: The Extents and Limits of Multilateralism*, Nueva York, Cambridge University Press, 2010. En la película, el juego es un poco diferente: los adolescentes conducen sus coches hacia un precipicio y cada uno intenta ser el segundo en saltar.

30. La impulsividad como una táctica paradójica: Frank, *Passions within Reason: The Strategic Role of the Emotions*, Nueva York, W. W. Norton, 1988; véase también Pinker, *How the Mind Works, op. cit.*, cap. 6.

31. Sagan y Suri, «The madman nuclear alert: Secrecy, signaling, and safety in October 1969», *International Security*, 27, 2003, págs. 150-183.

32. El amor loco como táctica paradójica: Frank, *Passions within Reason..., op. cit*.; Pinker, *How the Mind..., op. cit*., cap. 7, «Fools for love».

33. Novela de Dashiell Hammett, guion de John Huston.

34. Tetlock, «Thinking the unthinkable: Sacred values and taboo cognitions», *Trends in Cognitive Sciences*, 7, 2003, págs. 320-324, <https://doi.org/10.1016/S1364-6613(03)00135-9>; Tetlock, Kristel, Elson, Green y Lerner, «The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals», *Journal of Personality and Social Psychology*, 78, 2000, págs. 853-870, <https://doi.org/10.1037/0022-3514.78.5.853>.

35. Satel, *When Altruism Isn't Enough: The Case for Compensating Kidney Donors*, Washington, D. C., AEI Press, 2008.

36. Por ejemplo, Block, *Defending the Undefendable*, Auburn (AL), Ludwig von Mises Institute, 2018 [1976] (trad. cast.: *Defendiendo lo indefendible*, Madrid, Unión Editorial, 2012).

37. Sobre la reformulación de los intercambios tabúes, Tetlock, «Think ing the unthinkable...», art. cit.; Tetlock, Kristel, Elson, Green y Lerner, «The psychology of the unthinkable...», art. cit.; Zelizer, *The Purchase of Intimacy*, Princeton (NJ), Princeton University Press, 2005 (trad. cast.: *La negociación de la intimidad*, México, Fondo de Cultura Económica, 2010).

38. Hume, *A Treatise of Human Nature, op. cit*., libro III, sección III, «Of the influencing motives of the will». Sobre la filosofía moral de Hume, Cohon, «Hume’s moral...», art. cit.

39. Rachels y Rachels, *The Elements of Moral Philosophy*, Columbus (OH), McGraw-Hill, 2010 (trad. cast.: *Introducción a la filosofía moral*, Madrid, FCE, 2011).

40. Stoppard, *Jumpers...*, *op. cit.*, pág. 39.

41. Gould, *Rocks of Ages: Science and Religion in the Fullness of Life*, Nueva York, Ballantine, 1999 (trad. cast.: *Ciencia versus religión: un falso conflicto*, Barcelona, Booket, 2012).

42. Platón, «Euthyphro», en J. M. Cooper (comp.), *Plato: Five Dialogues – Euthyphro*, *Apology, Crito, Meno, Phaedo*, Indianápolis, Hackett, 2002 [399-390 a. C.] (trad. cast.: «Eutifrón», en *Diálogos*, vol. 1, Biblioteca Clásica Gredos, Barcelona, 2019). La filosofía moral de Platón cobra vida en R. Goldstein, *Plato at the Googleplex: Why Philosophy Won't Go Away*, Nueva York, Pantheon, 2013.

43. Sobre cómo Dios ordena el sacrificio de niños, Pinker, *The Better Angels...*, *op. cit*., cap. 1.

44. «Tampoco es contrario a la razón preferir un bien pequeño, aunque lo reconozca como menor, a otro mayor, y tener una afección más ardiente por el primero que por el segundo.»

45. Sobre la moralidad como imparcialidad: De Lazari-Radek y Singer, «The objectivity of ethics and the unity of practical reason», *Ethics*, 123, 2012, págs. 9-31, <https://doi.org/10.1086/667837>; Goldstein, *Betraying Spinoza: The Renegade Jew Who Gave Us Modernity*, Nueva York, Nextbook/Shocken, 2006; Greene, *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*, Nueva York, Penguin, 2013; Nagel, *The Possibility of Altruism*, Princeton (NJ), Princeton University Press, 1970 (trad. cast.: *La posibilidad del altruismo*, México, Fondo de Cultura Económica, 2004); Railton, «Moral realism», *Philosophical Review*, 95, 1986, págs. 163-207, <https://doi.org/10.2307/2185589>; Singer, *The Expanding Circle: Ethics and Sociobiology*, Princeton (NJ), Princeton University Press, 2011 [1981].

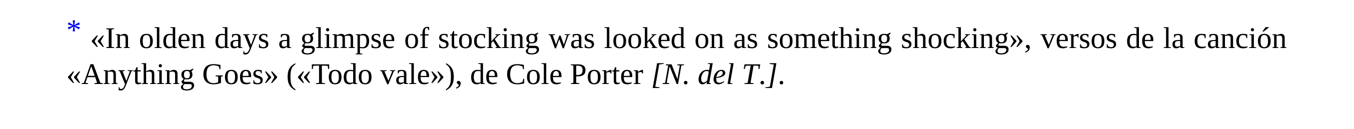
46. Terry, *Golden Rules and Silver Rules of Humanity: Universal Wisdom of Civilization*, Berkeley (CA), AuthorHouse, 2008.

47. Sobre interés propio, sociabilidad y racionalidad como condiciones suficientes para la moralidad: Pinker, *Enlightenment Now...*, *op. cit*., págs. 412-415. La moralidad como estrategia en los juegos de suma positiva: Pinker, *The Better Angels...*, *op. cit*., págs. 689-692.

48. Chomsky, *Language and Mind*, Nueva York, Cambridge University Press, 2006 [1972] (trad. cast.: *El lenguaje y el entendimiento*, Barcelona, Seix Barral, 1986); Pinker, *The Language Instinct*, Nueva York, HarperCollins, 2007 [1994] (trad. cast.: *El instinto del lenguaje: cómo la mente construye el lenguaje*, Madrid, Alianza, 2012), cap. 4.



* Respectivamente, «*a man needs a little madness or else he never dares cut the rope and be free*», «*stop making sense*» y «*let’s go crazy*» *[N. del T.]*.


* «In olden days a glimpse of stocking was looked on as something shocking», versos de la canción «Anything Goes» («Todo vale»), de Cole Porter *[N. del T.]*.

1. Eliot, G., *Essays of George Eliot*, Filadelfia, Routledge, 2017 [1883], págs. 257-258.

2. Leibniz, «On universal synthesis and analysis, or the art of discovery and judgment», en L. E. Loemker (comp.), *Philosophical Papers and Letters*, Nueva York, Springer, 1989 [1679].

3. Sobre las introducciones accesibles a la lógica, McCawley, *Everything That Linguists Have Always Wanted to Know about Logic, but Were Ashamed to Ask*, Chicago, University of Chicago Press, 1993 (2.ª ed.); Priest, *Logic: A Very Short Introduction*, Nueva York, Oxford University Press, 2017 (2.ª ed.) (trad. cast.: *Una brevísima introducción a la lógica*, México, Océano, 2006); Warburton, *Thinking from A to Z*, Nueva York, Routledge, 2007 (3.ª ed.) (trad. cast.: *Pensar de la A a la Z*, Barcelona, Gedisa, 2016).

4. Basado en Carroll, «Symbolic logic», en W. W. Bartley (comp.), *Lewis Carroll's Symbolic Logic*, Nueva York, Clarkson Potter, 1977 [1896], libro II, cap. 3, §2, ejemplo (4), pág. 72.

5. Donaldson, Doubleday, Hefferman, *et al.*, «Are talking heads...», art. cit.

6. Términos lógicos en la lógica frente a la conversación: Grice, «Logic and conversation», art. cit.; Pinker, *The Language...*, *op. cit*., caps. 2 y 8.

7. Emerson, *Self-reliance and Other Essays*, Nueva York, Dover, 1993 [1841] (trad. cast.: *Confianza en uno mismo*, Madrid, Gadir, 2009).

8. Liberman, «If P, so why not Q?», *Language Log*, 5 de agosto de 2004, <http://itre.cis.upenn.edu/~myl/languagelog/archives/001314.html>.

9. McCawley, *Everything That Linguists...*, *op. cit*.

10. Del sitio web de Yang 2020, recuperado el 6 de febrero de 2020.

11. Curtis, «The Fallacy Files taxonomy of logical fallacies», 2020, <https://www.fallacyfiles.org/taxonnew.htm>; Richardson, Smith, Mead en y Flip Creative, «Thou shalt not commit logical fallacies», 2020, <https://yourlogicalfallacyis.com/>; Warburton, *Thinking...*, *op. cit.*; véase también el artículo de Wikipedia «List of fallacies», <https://en.wikipedia.org/wiki/List_of_fallacies>.

12. Mercier y Sperber, «Why do humans...», art. cit.; para una crítica, véase Norman, «Why we reason: Intention-alignment and the genesis of human rationality», *Biology and Philosophy*, 31, 2016, págs. 685-704, <https://doi.org/10.1007/s10539-016-9532-4>.

13. Friedersdorf, «Why can't people hear what Jordan Peterson is actually saying?», *The Atlantic*, 22 de enero de 2018, <https://www.theatlantic.com/politics/archive/2018/01/putting-monsterpaint-onjordan-peterson/550859/>.

14. Shackel, «Motte and bailey doctrines», 2014, <https://blog.practicalethics.ox.ac.uk/2014/09/motte-and-bailey-doctrines/>.

15. Russell, «Letter to Mr. Major», en B. Feinberg y R. Kasrils (comps.), *Dear Bertrand Russell: A Selection of his Correspondence with the General Public 1950-1968*, Londres, Allen & Unwin, 1969.

16. Basterfield, Lilienfeld, Bowes y Costello, «The Nobel disease: when intelligence fails to protect against irrationality», *Skeptical Inquirer*, mayo de 2020, <https://skepticalinquirer.org/2020/05/the-nobel-disease-whenintelligence-fails-to-protect-against-irrationality/>.

17. Un dicho común basado libremente en un pasaje de *Un enemigo del pueblo*, de Henrik Ibsen: «La mayoría nunca tiene la razón de su lado [...]. La mayoría tiene la fuerza de su lado, por desgracia; pero no tiene la razón».

18. Proctor, *The Nazi War on Cancer*, Princeton (NJ), Princeton University Press, 2000.

19. Para la discusión de un ejemplo, véase Paresky, Haidt, Strossen y Pinker, «*The New York Times* surrendered to an outrage mob. Journalism will suffer for it», *Politico*, 14 de mayo de 2020, <https://www.politico.com/news/magazine/2020/05/14/bret-stephens-new-york-times-outragebacklash-256494>.

20. Haidt, *The Righteous Mind: Why Good People Are Divided by Politics and Religion*, Nueva York, Pantheon, 2012 (trad. cast.: *La mente de los justos: por qué la política y la religión dividen a la gente sensata*, Barcelona, Deusto, 2019).

21. La historia se encuentra en muchos libros de texto y suele atribuirse a Francis Bacon en 1592, pero su auténtica fuente, incluso como parodia, es oscura, y probablemente date de principios del siglo XX; véase Simanek, D., «Horse’s Teeth», 1999, <https://www.lockhaven.edu/~dsimanek/horse.htm>.

22. Sobre la racionalidad ecológica, Gigerenzer, «Ecological intelligence...», art. cit.; Pinker, *How the Mind...*, *op. cit*., págs. 302-306; Tooby y Cosmides, «Ecological rationality and the multimodular mind: Grounding normative theories in adaptive problems», en K. I. Manktelow y D. E. Over (comps.), *Rationality: Psychological and Philosophical Perspectives*, Londres, Routledge, 1993.

23. Cosmides, «The logic of social exchange...», art. cit.; Fiddick, Cosmides y Tooby, «No interpretation...», art. cit.

24. Weber, *Economy and Society: A New Translation*, Cambridge (MA), Harvard University Press, 2019 [1922] (trad. cast.: *Economía y sociedad: esbozo de sociología comprensiva*, Madrid, Fondo de Cultura Económica, 2002).

25. Cole, Gay, Glick y Sharp, *The Cultural Context of Learning and Thinking*, Nueva York, Basic Books, 1971, págs. 187-188; véase también Scribner y Cole, «Cognitive consequences of formal and informal education», *Science*, 182, 1973, págs. 553-559, <https://doi.org/10.1126/science.182.4112.553>.

26. Norenzayan, Smith, Kim y Nisbett, «Cultural preferences for formal versus intuitive reasoning», *Cognitive Science*, 26, 2002, págs. 653-684.

27. Wittgenstein, *Philosophical Investigations*, Nueva York, Macmillan, 1953 (trad. cast.: *Investigaciones filosóficas*, Madrid, Trotta, 2017).

28. No todos los filósofos están de acuerdo: Bernard Suits, *The Grasshopper: Games, Life, and Utopia*, Peterborough, Broadview Press, 2014 [1978], define un juego como «el intento voluntario de superar obstáculos innecesarios». Véase también McGinn, *Truth by Analysis: Games, Names, and Philosophy*, Nueva York, Oxford University Press, 2012, cap. 2.

29. Pinker, *How the Mind...*, *op. cit.*, págs. 306-313; Pinker, *Words and Rules: The Ingredients of Language*, Nueva York, HarperCollins, 2011 [1999], cap. 10; Pinker y Prince, «The nature of human concepts: Evidence from an unusual source», en S. Pinker (comp.), *Language, Cognition, and Human Nature: Selected Articles*, Nueva York, Oxford University Press, 2013; Rosch, «Principles of categorization», en E. Rosch y B. B. Lloyd (comps.), *Cognition and Categorization*, Hillsdale (NJ), Erlbaum, 1978.

30. Armstrong, Gleitman y Gleitman, «What some concepts might not be», *Cognition*, 13, 1983, págs. 263-308, <https://doi.org/10.1016/00100277(83)90012-4>; Pinker, *Words and Rules..., op. cit*., cap. 10; Pinker y Prince, «The nature of human concepts...», art. cit.

31. Goodfellow, Bengio y Courville, *Deep Learning*, Cambridge (MA), MIT Press, 2016; Rumelhart, McClelland y PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1: *Foundations*, Cambridge (MA), MIT Press, 1986; Aggarwal, *Neural Networks and Deep Learning*, Nueva York, Springer, 2018. Para las visiones críticas, véanse Marcus y Davis, *Rebooting AI..., op. cit.*; Pearl y Mackenzie, *The Book of Why: The New Science of Cause and Effect*, Nueva York, Basic Books, 2018 (trad. cast.: *El libro del porqué: la nueva ciencia de la causa y el efecto*, Barcelona, Pasado y Presente, 2020); Pinker, *Words and Rules..., op. cit.*; Pinker y Mehler (comps.), *Connections and Symbols*, Cambridge (MA), MIT Press, 1988.

32. Rumelhart, Hinton y Williams, «Learning representations by back-propagating errors», *Nature*, 323, 1986, págs. 533-536, <https://doi.org/10.1038/323533a0>; Aggarwal, *Neural Networks..., op. cit*.; Goodfellow, Bengio y Courville, *Deep Learning..., op. cit*.

33. Lewis-Kraus, «The great AI awakening», *The New York Times Magazine*, pág. 12, 14 de diciembre de 2016, <https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html>.

34. La palabra *algoritmo* se reservaba originalmente para esas fórmulas, que contrastaban con la *heurística* o reglas generales. Pero, en el lenguaje corriente actual, se emplea para todos los sistemas de inteligencia artificial, incluidos los basados en las redes neuronales.

35. Marcus y Davis, *Rebooting AI...*, *op. cit*.

36. Kissinger, «How the Enlightenment ends», *The Atlantic*, junio de 2018, <https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/>.

37. Lake, Ullman, Tenenbaum y Gershman, «Building machines that learn and think like people», *Behavioral and Brain Sciences*, 39, 2017, págs. 1-101, <https://doi.org/10.1017/S0140525X16001837>; Marcus, «The deepest problem with deep learning», *Medium*, 1 de diciembre de 2018, <https://medium.com/@GaryMarcus/the-deepest-problem-with-deeplearning-91c5991f5695>; Marcus y Davis, *Rebooting AI*..., *op. cit*.; Pearl y Mackenzie, *The Book of Why*..., *op. cit*.

38. Ashby, Alfonso-Reese, Turken y Waldron, «A neuropsychological theory of multiple systems in category learning», *Psychological Review*, 105, 1998, págs. 442-481, <https://doi.org/10.1037/0033-295X.105.3.442>; Evans, «Dual-process theories of deductive reasoning: Facts and fallacies», en K. J. Holyoak y R. G. Morrison (comps.), *The Oxford Handbook of Thinking and Reasoning*, Oxford, Oxford University Press, 2012; Kahneman, *Thinking, Fast and Slow, op. cit.*; Marcus, «Two kinds of representation», en E. Dietrich y A. B. Markman (comps.), *Cognitive Dynamics: Conceptual and Representational Change in Humans and Machines*, Mahwah (NJ), Erlbaum, 2000; Pinker, *Words and Rules..., op. cit.*; Pinker y Prince, «The nature of human concepts...», art. cit., págs. 3-22, <https://doi.org/10.1037/00332909.119.1.3>.

39. Pinker, *Words and Rules...*, *op. cit.*, cap. 10; Pinker y Prince, «The nature of human concepts...», art. cit.

* Bill Clinton y Monica Lewinsky *[N. del T.]*.

1. Carta a la señorita Sophia Thrale, 24 de julio de 1783, en Johnson, *The Letters of Samuel Johnson with Mrs. Thrale's Genuine Letters to Him*, Nueva York, Oxford University Press, 1963.

2. *Bartlett's Familiar Quotations*. La cita no conduce a una fuente primaria, pero probablemente fuese una carta a Max Born de 1926. Una variante aparece en una carta a Cornelius Lanczos, citada en Dukas, H., y Hoffman, B. (comps.), *Albert Einstein, The Human Side: New Glimpses from his Archives*, Princeton (NJ), Princeton University Press, 1981, y pueden encontrarse otras tres en la entrada de Einstein en Wikiquote, <https://en.wikiquote.org/wiki/Albert_Einstein>.

3. Eagle, «Chance versus randomness», en E. N. Zalta (comp.), *The Stanford Encyclopedia of Philosophy*, 2019, <https://plato.stanford.edu/entries/chance-randomness/>; la aleatoriedad como incompresibilidad, habitualmente denominada complejidad de Kolmogorov, se discute en la sección 2.2.1.

4. Millenson, «An inexpensive Geiger gate for controlling probabilities of events», *Journal of the Experimental Analysis of Behavior*, 8, 1965, págs. 345-346.

5. Póster de la gravedad: <http://www.mooneyart.com/gravity/historyof_01.html>.

6. Gigerenzer, Hertwig, Broek, Fasolo y Katsikopoulos, «"A 30 % chance of rain tomorrow": How does the public understand probabilistic weather forecasts?», *Risk Analysis: An International Journal*, 25, 2005, págs. 623-629, <https://doi.org/10.1111/j.1539-6924.2005.00608.x>.

7. Citado en Bell, *The Development of Mathematics*, Nueva York, McGraw Hill, 1947 (2.ª ed.) (trad. cast.: *Historia de las matemáticas*, México, FCE, 1949).

8. Interpretaciones de la probabilidad: Gigerenzer, «The evolution of statistical thinking», en G. Gigerenzer (comp.), *Rationality for Mortals: How People Cope with Uncertainty*, Nueva York, Oxford University Press, 2008; Gigerenzer, Swijtink, Porter *et al*., *The Empire of Chance...*, *op. cit*.; Hájek, «Interpretations of probability», en E. N. Zalta (comp.), *The Stanford Encyclopedia of Philosophy*, 2019, <https://plato.stanford.edu/archives/fall2019/entries/probability-interpret/>; Savage, *The Foundations of Statistics*, Nueva York, Wiley, 1954.

9. Citado en Gigerenzer, «How to make cognitive illusions disappear: Beyond “heuristics and biases”», *European Review of Social Psychology*, 2, 1991, págs. 83-115, <https://doi.org/10.1080/14792779143000033>, pág. 8.

10. Gigerenzer, *Rationality for Mortals..., op. cit*.

11. Tversky y Kahneman, «Availability: A heuristic for judging frequency and probability», *Cognitive Psychology*, 5, 1973, págs. 207-232, <https://doi.org/10.1016/0010-0285(73)90033-9>.

12. Gigerenzer, *Rationality for Mortals...*, *op. cit*.

13. Combs y Slovic, «Newspaper coverage of causes of death», *Journalism Quarterly*, 56, 1979, págs. 837-849; Ropeik, *How Risky Is It, Really? Why Our Fears Don't Always Match the Facts*, Nueva York, McGraw-Hill, 2010; Slovic, «Perception of risk», *Science*, 236, 1987, págs. 280-285, <https://doi.org/10.1126/science.3563507>.

14. McCarthy, «Americans still greatly overestimate U.S. gay population», *Gallup*, 2019, <https://news.gallup.com/poll/259571/americansgreatly-overestimate-gay-population.aspx>.

15. Duffy, *The Perils of Perception: Why We're Wrong about Nearly Everything*, Londres, Atlantic Books, 2018; véase también Ropeik, *How Risky Is It...*, *op. cit.*; Slovic, «Perception of risk», *Science*, 236, 1987, págs. 280-285, <https://doi.org/10.1126/science.3563507>.

16. Cifras de 2014-2015, citadas en Pinker, *Enlightenment Now...*, *op. cit*., tabla 13-1, pág. 192. Véase también Ritchie, «Causes of death», art. cit.; Roth, Abate, Abate, *et al*., «Global, regional, and national agesexspecific mortality for 282 causes of death in 195 countries and territories, 19802017: a systematic analysis for the Global Burden of Disease Study 2017», *The Lancet*, 392, 2018, págs. 1736-1788, <https://doi.org/10.1016/S01406736(18)32203-7>.

17. Savage, Comparing the fatality risks in United States transportation across modes and over time», *Research in Transportation Economics*, 43, 2013, págs. 9-22, <https://doi.org/10.1016/j.retrec.2012.12.011>, tabla 2. La cifra corresponde a la aviación comercial en Estados Unidos.

18. Gigerenzer, «Out of the frying pan into the fire: Behavioral reactions to terrorist attacks», *Risk Analysis*, 26, 2006, págs. 347-351, <https://doi.org/10.1111/j.1539-6924.2006.00753.x>.

19. «Mack the Knife» («Mackie el Navaja»), letra de Bertolt Brecht, de *La ópera de los tres centavos*.

20. Tiburones de Cape Cod: Sherman, «The shark attack that changed Cape Cod forever», *Boston Magazine*, 14 de mayo de 2019, <https://www.bostonmagazine.com/news/2019/05/14/cape-cod-sharks/>. Muertes por accidentes de tráfico en Cape Cod: Nolan, Bremer, Tupper *et al.*, *Barnstable County High Crash Locations: Cape Cod Commission*, 2019, <https://www.capecodcommission.org/resource-library/file/?url=/dept/commission/team/tr/Reference/Safety-General/Top50CrashLocs_2018Final.pdf>.

21. Caldeira, Emanuel, Hansen y Wigley, «Top climate change scientists' letter to policy influencers», CNN, 3 de noviembre de 2013, <https://www.cnn.com/2013/11/03/world/nuclear-energy-climate-change-scientists-letter/index.html>. Véanse también Goldstein y Qvist, *A Bright Future: How Some Countries Have Solved Climate Change and the Rest Can Follow*, Nueva York, PublicAffairs, 2019; Goldstein, Qvist y Pinker, «Nuclear power can save the world», *The New York Times*, 6 de abril de 2019, <https://www.nytimes.com/2019/04/06/opinion/sunday/climate-changenuclear-power.html>.

22. Sobre la energía nuclear versus el carbón: Goldstein y Qvist, *A Bright Future...*, *op. cit.*; Goldstein, Qvist y Pinker, «Nuclear power...», art. cit. Sobre cómo el carbón mata: Lockwood, Welker-Hood, Rauch y Gottlieb, *Coal's Assault on Human Health: A Report from Physicians for Social Responsibility*, <https://www.psr.org/blog/resource/coals-assault-on-humanhealth/>, 2009. Sobre la nuclear reemplazada por el carbón, Jarvis, Deschenes y Jha, *The Private and External Costs of Germany's Nuclear Phase-out*, 2019, <https://haas.berkeley.edu/wp-content/uploads/WP304.pdf>. Incluso si aceptáramos las recientes afirmaciones de que las autoridades encubrieron miles de muertes en Chernóbil, el número de víctimas de sesenta años de energía nuclear equivaldría aún aproximadamente a un mes de muertes relacionadas con el carbón.

23. Ropeik, *How Risky Is It...*, *op. cit*.; Slovic, «Perception...», art. cit.

24. Pinker, *Enlightenment Now..., op. cit*.; tabla 13.1, pág. 192; Mueller, *Overblown: How Politicians and the Terrorism Industry Inflate National Security Threats, and Why We Believe Them*, Nueva York, Free Press, 2006.

25. Walker, Petulla, Fowler *et al*., «10 years. 180 school shootings. 356 victims», CNN, julio de 2019, <https://www.cnn.com/interactive/2019/07/us/ten-years-of-school-shootings-trnd/>.

26. Los promedios corresponden a 20152019. El número de asesinatos policiales, a Tate, Jenkins, Rich *et al*., «Fatal force», *The Washington Post*, consultado el 14 de octubre de 2020, <https://www.washingtonpost.com/graphics/investigations/policeshootingsdatabase/>. En número de homicidios, en FBI, FBI, «Crime in the United States, expanded homicide data table 1», 2019, <https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s2019/tables/expanded-homcide-data-table-1.xls> y años anteriores.

27. Schelling, *The Strategy...*, *op. cit.*, pág. 90; véase también Tooby, Cosmides y Price, «Cognitive adaptations for *n*-person exchange: The evolutionary roots of organizational behavior», *Managerial and Decision Economics*, 27, 2006, págs. 103-129, <https://doi.org/10.1002/mde.1287>. Pearl Harbor y el 11S como ultrajes públicos: Mueller, *Overblown...*, *op. cit.*

28. Chwe, *Rational Ritual: Culture, Coordination, and Common Knowledge*, Princeton (NJ), Princeton University Press, 2001; De Freitas, Thomas, DeScioli y Pinker, «Common knowledge, coordination, and strategic mentalizing in human social life», *Proceedings of the National Academy of Sciences*, 116, 2019, págs. 13751-13758, <https://doi.org/10.1073/pnas.1905518116>; Schelling, *The Strategy...*, *op. cit*.

29. Baumeister, Stillwell y Wotman, «Victim and perpetrator accounts of interpersonal conflict: Autibiographical narratives about anger», *Journal of Personality and Social Psychology*, 59, 1990, págs. 994-1005, <https://doi.org/10.1037/0022-3514.59.5.994>.

30. Hostilidad de los datos sobre los ultrajes públicos en los casos de Pearl Harbor y el 11S, Mueller, *Overblown...*, *op. cit*.; en el caso del asesinato de George Floyd, Blackwell, «Black Lives Matter and the mechanics of conformity», *Quillette*, 17 de septiembre de 2020, <https://quillettte.com/2020/09/17/black-lives-matter-and-the-mechanics-of-conformity/>.

31. Popularizado por Rahm Emanuel, jefe de Gabinete de Obama, pero utilizado anteriormente por el antropólogo Luther Gerlach. Gracias a Fred Shapiro, editor de *The Yale Book of Quotations*.

32. Para un extenso argumento de esta clase a propósito del terrorismo, véase Mueller, *Overblown...*, *op. cit*.

33. Véase <https://twitter.com/MaxCRoser/status/919921745464905728?s=20>.

34. McCarthy, «More Americans say crime is rising in U.S.», *Gallup*, 22 de octubre de 2015, <https://news.gallup.com/poll/186308/americanssay-crime-rising.aspx>.

35. Rosling, *Factfulness: Ten Reasons We're Wrong about the World and Why Things Are Better Than You Think*, Nueva York, Flatiron, 2019 (trad. cast.: *Factfulness: diez razones por las que estamos equivocados sobre el mundo y por qué las cosas están mejor de lo que piensas*, Barcelona, Deusto, 2019).

36. Sobre los medios de comunicación impulsados por las crisis y por el cinismo político, Bornstein y Rosenberg, «When reportage turns to cynicism», *The New York Times*, 14 de noviembre de 2016, <https://www.nytimes.com/2016/11/15/opinion/when-reportage-turns-to-cynicism.html>.

37. Lankford y Madfis, «Don’t name them, don’t show them, but report everything else: A pragmatic proposal for denying mass killers the attention they seek and deterring future offenders», *American Behavioral Scientist*, 62, 2018, págs. 260-279, <https://doi.org/10.1177/0002764217730854>.

38. Véase <www.ourworldindata.org>.

39. Tomado de Paulos, *Innumeracy: Mathematical Illiteracy and Its Consequences*, Nueva York, MacMillan, 1988 (trad. cast.: *El hombre anumérico: el analfabetismo matemático y sus consecuencias*, Barcelona, Tusquets, 2016).

40. Edwards, «Is the Pope an alien?», *Nature*, 382, 1996, pág. 202, <https://doi.org/10.1038/382202b0>.

41. Muchos libros explican la probabilidad y sus escollos, incluidos Paulos, *Innumeracy...*, *op. cit.*; Hastie y Dawes, *Rational Choice...*, *op. cit.*; Mlodinow, *The Drunkard's Walk: How Randomness Rules Our Lives*, Nueva York, Vintage, 2009 (trad. cast.: *El andar del borracho: cómo el azar gobierna nuestras vidas*, Barcelona, Crítica, 2010); Schneps y Colmez, *Math on Trial: How Numbers Get Used and Abused in the Courtoom*, Nueva York, Basic Books, 2013.

42. Batt, *Stolen Innocence: A Mother's Fight for Justice – The Authorized Story of Sally Clard*, Londres, Ebury Press, 2004; Schneps y Colmez, *Math on Trial...*, *op. cit*.

43. Texas versus Pensilvania, 2020. Moción: <https://www.texasattorneygeneral.gov/sites/default/files/images/admin/2020/Press/SCOTUSFiling.pdf>. Expediente: <https://www.supremecourt.gov/docket/docketfiles/html/public/22O155.html>. Análisis: Bump, «Trump's effort to steal the election comes down to some utterly ridiculous statistical claims», *The Washington Post*, 9 de diciembre de 2020, <https://www.washingtonpost.com/politics/2020/12/09/trumps-effort-steal-election-comes-down-someutterly-ridiculous-statistical-claims/>.

44. Gilovich, Vallone y Tversky, «The hot hand in basketball: On the misperception of random sequences», *Cognitive Psychology*, 17, 1985, págs. 295-314, <https://doi.org/10.1016/0010-0285(85)90010-6>.

45. Miller y Sanjurjo, «Surprised by the hot hand fallacy? A truth in the law of small numbers», *Econometrica*, 86, 2018, págs. 2019-2047, <https://doi.org/10.3982/ECTA14943>, Gigerenzer, «The Bias Bias in behavioral economics», *Review of Behavioral Economics*, 5, 2018, págs. 303-336, <https://doi.org/10.1561/105.00000092>.

46. Pinker, *The Better Angels...*, *op. cit*., págs. 202-207.

47. Véase <https://xkcd.com/795/>.

48. Krämer y Gigerenzer, «How to confuse with statistics or: the use and misuse of conditional probabilities», *Statistical Science*, 20, 2005, págs. 223-230, <https://doi.org/10.1214/088342305000000029>.

49. *Ibidem*; Miller y Sanjurjo, «Surprised by the hot hand fallacy...», art. cit.; Miller y Sanjurjo, «A bridge from Monty Hall to the hot hand: The principle of restricted choice», *Journal of Economic Perspectives*, 33, 2019, págs. 144-162, <https://doi.org/10.1257/jep.33.3.144>.

50. Véase <https://www.youtube.com/watch?v=DBSAeqdcZAM>.

51. La crítica de Scarry se describe en Rosen, «The bloods and the crits», *New Republic*, 9 de diciembre de 1996, <https://newrepublic.com/article/74070/the-bloods-and-the-crits>; véase también Good, «When batterer becomes murderer», *Nature*, 381, 1996, pág. 481, <https://doi.org/10.1038/381481a0>.

52. Krämer y Gigerenzer, «How to confuse.», art. cit.

53. *Ibidem*; Schneps y Colmez, *Math on Trial...*, *op. cit*.

54. Artículo en Johnson, Tress, Burkel, *et al*., «Officer characteristics and racial disparities in fatal officerinvolved shootings», *Proceedings of the National Academy of Sciences*, 116, 2019, págs. 15877-15882, <https://doi.org/10.1073/pnas.1903856116>. Crítica en Knox y Mummolo, «Making inferences about racial disparities in police violence», *Proceedings of the National Academy of Sciences*, 117, 2020, págs. 1261-1262, <https://doi.org/10.1073/pnas.1919418117>. Respuesta en Johnson y Cesario, «Reply to Knox and Mummolo and Schimmack and Carlsson: Controlling for crime and population rates», *Proceedings of the National Academy of Sciences*, 117, 2020, págs. 1264-1265, <https://doi.org/10.1073/pnas.1920184117>. Retractación en Cesario y Johnson, «Statement on the retraction of "Officer characteristics and racial disparities in fatal officerinvolved shootings"», 2020, <https://doi.org/10.31234/osf.io/dj57k>.

55. Edwards, «Is the Pope...», art. cit.

56. Mlodinow, *The Drunkard's Walk...*, *op. cit*.; Paulos, *Innumeracy...*, *op. cit*.

57. Fabrikant, «Humbler, after a streak of magic», *The New York Times*, 11 de mayo de 2008, <https://www.nytimes.com/2008/05/11/business/11bill.html>; Mlodinow, *The Drunkard's Walk...*, *op. cit.*; Serwer, «The greatest money manager of our time», *CCN Money*, 15 de noviembre de 2006, <https://money.cnn.com/magazines/fortune_archive/2006/11/27/8394343/index.htm>.

58. Gardner, «Why the long arm of coincidence is usually not as long as it seems», *Scientific American*, 227, 1972.

59. Open Science Collaboration, «Estimating the reproductibility of psychological science», *Science*, 349, 2015, <https://doi.org/10.1126/science.aac4716>; Gigerenzer, «Statistical rituals: The replication delusion and how we got there», *Advances in Methods and Practices in Psychological Science*, 1, 2018, págs. 198-218, <https://doi.org/10.1177/2515245918771329>; Ioannidis, «Why most published research findings are false», *PloS Medicine*, 2, 2005, e124, <https://doi.org/10.1371/journal.pmed.0020124>; Pashler y Wagenmakers, «Editors' introduction to the special section on replicability in psychological science: A crisis of confidence?», *Perspectives on Psychological Science*, 7, 2012, págs. 528-530, <https://doi.org/10.1177/1745691612465253>.

60. Ioannidis, «Why most published...», art. cit.; Simmons, Nelson y Simonsohn, «Falsepositive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant», *Psychological Science*, 22, 2011, págs. 1359-1366, <https://doi.org/10.1177/0956797611417632>. «El jardín de los senderos que se bifurcan» fue acuñado por el estadístico Andrew Gelman (Gelman y Loken, «The statistical crisis in science», *American Scientist*, 102, 2014, págs. 460-465).

61. El psicólogo cognitivo Michael Corballis.

62. Por ejemplo, los Registros OSF del Center for Open Science, <https://osf.io/prereg/>.

63. Feller, *An Introduction to Probability Theory and Its Applications*, Nueva York, Wiley, 1968 (trad. cast.: *Introducción a la teoría de probabilidades y sus aplicaciones*, México, Limusa, 1978); véase Pinker, *The Better Angels*..., *op. cit*., págs. 202-207.

64. Kahneman y Tversky, «Subjective probability: A judgment of representativeness», *Cognitive Psychology*, 3, 1972, 3, págs. 430-454, <https://doi.org/10.1016/0010-0285(72)90016-3>. Mostrados originalmente por Feller, *An Introduction to Probability...*, *op. cit*.

65. Gould, «The streak...», art. cit.

* Evocación del verso *«oh, I have slipped the surly bonds of earth»* con el que comienza el poema «High Flight», del poeta y piloto de combate John Gillespie Magee *[N. del T.]*.

* *«When that shark bites with his teeth, babe, scarlet billows start to spread» [N. del T.].*

* Míchigan y Wisconsin respectivamente *[N. de T.]*.

1. Sobre la *rationality community* («comunidad de racionalidad»), Caplan, «What’s wrong with the rationality community», *EconLog*, 4 de abril de 2017, <https://www.econlib.org/archives/2017/04/whats_wrong_wit_22.html>; Chivers, *The AI Does Not Hate You: Superintelligence, Rationality and the Race to Save the World*, Londres, Weidenfeld & Nicolson, 2019; Raemon, «What exactly is the “Rationality Community”?», *LessWrong*, 9 de abril de 2017, <https://www.lesswrong.com/posts/s8yvtCbbZW2S4WnhE/what-exactly-is-the-rationality-community>. Entre sus miembros prominentes figuran Julia Galef, de *Rationally Speaking* (<https://juliagalef.com>); Scott Alexander, de *Slate Star Codex* (<https://slatestarcodex.com/>); Scott Aaronson, de *Shtetl-Optimized* (<https://www.scottaaronson.com/blog/>); Robin Hanson, de *Overcoming Bias* (<https://www.overcomingbias.com/>); y Eliezer Yudkowsky, que creó *Less Wrong* (<https://www.lesswrong.com/>).

2. Arbital, «Bayes’ rule», 2020, <https://arbital.com/p/bayes_rule/?l=1zq>.

3. Gigerenzer, «What are natural frequencies?», *BMJ*, 343, 2011, d6386. <https://doi.org/10.1136/bmj.d6386>.

4. Más exactamente, P(datos | hipótesis) es *proporcional* a la verosimilitud. El término *verosimilitud* tiene significados técnicos ligeramente diferentes en las distintas subcomunidades estadísticas; este es el que se emplea habitualmente en las discusiones del razonamiento bayesiano.

5. Kahneman y Tversky, «Subjective probability...», art. cit.; Tversky y Kahneman, «Judgment under uncertainty...», art. cit.

6. «En su evaluación de las evidencias, aparentemente el hombre no es un bayesiano conservador: no es bayesiano en absoluto», Kahneman y Tversky, «Belief in the law of small numbers», *Psychological Bulletin*, 76, 1971, págs. 105-110, <https://doi.org/10.1037/h0031322>, pág. 450.

7. Tversky y Kahneman, «Evidential impact of base rates», en D. Kahneman, P. Slovic y A. Tversky (comps.), *Judgment under Uncertainty: Heuristics and Biases*, Nueva York, Cambridge University Press, 1982.

8. Hastie y Dawes, *Rational Choice...*, *op. cit*.

9. Tversky y Kahneman, «Judgment under uncertainty...», art. cit., págs. 1124-1131, <https://doi.org/10.1126/science.185-4157.1124>.

10. Oído casualmente; no encuentro ninguna versión impresa.

11. Sobre Hume, Bayes y los milagros, Earman, «Bayes, Hume, Price, and miracles», *Proceedings of the British Academy*, 113, 2002, págs. 91-109.

12. Hume, *An Enquiry Concerning Human Understanding*, Nueva York, Oxford University Press, 1999 [1748] (trad. cast.: *Investigación sobre el entendimiento humano*, Madrid, Istmo, 2004), sección X, «Of miracles», parte 1, pág. 90.

13. *Ibidem*, pág. 91.

14. French, «Precognition studies and the curse of the failed replications», *The Guardian*, 15 de marzo de 2012, <http://www.theguardian.com/science/2012/mar/15/precognition-studies-curse-failed-replications>.

15. Carroll, *The Big Picture: On the Origins of Life, Meaning, and the Universe Itself*, Nueva York, Penguin Random House, 2016. Véase también Stenger, *Physics and Psychics: the Search for a World Beyond the Senses*, Búfalo, Nueva York, Prometheus, 1990.

16. Open Science Collaboration, «Estimating the reproductibility...», art. cit.; Pashler y Wagenmakers, «Editors' introduction...», art. cit.

17. Sobre la inefectividad de las industrias de la persuasión, en Mercier, *Not Born Yesterday: The Science of Who We Trust and What We Believe*, Princeton (NJ), Princeton University Press, 2020.

18. Ziman, *Reliable Knowledge: An Exploration of the Grounds for Belief in Science*, Nueva York, Cambridge University Press, 1978, pág. 40.

19. Tetlock y Gardner, *Superforecasting...*, *op. cit*.

20. Tetlock, «Thinking the unthinkable...», art. cit.; Tetlock, Kristel, Elson, Green y Lerner, «The psychology...», art. cit.

21. Descenso del fanatismo: Pinker, *Enlightenment Now...*, *op. cit.*, págs. 215-219; Charlesworth y Banaji, «Patterns of implicit and explicit attitudes: I. Long-term change and stability from 2007 to 2016», *Psychological Science*, 30, 2019, págs. 174-192, <https://doi.org/10.1177/0956797618813087>.

22. Política de tasas base en las ciencias sociales: Tetlock, «Political psychology or politicized psychology: Is the road to scientific hell paved with good moral intentions?», *Political Psychology*, 15, 1994, págs. 509-529, <https://doi.org/10.2307/3791569>.

23. Gigerenzer, «How to make cognitive illusions disappear: Beyond “heuristics and biases”», *European Review of Social Psychology*, 2, 1991, págs. 83-115, <https://doi.org/10.1080/14792779143000033>; «The Bias Bias...», art. cit.; Gigerenzer, Porter, Daston *et al*., *The Empire of Chance...*, *op.cit*.; véase también Cosmides y Tooby, «Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty», *Cognition*, 58, 1996, págs. 1-73, <https://doi.org/10.1016/0010-0277(95)00664-8>.

24. Burns, «At veterinary colleges, male students are in the minority», *American Veterinary Medical Association*, 15 de febrero de 2010, <https://www.avma.org/javma-news/2010-02-15/veterinary-colleges-male-students-are-minority>; Maines, «Why are women crowding into schools of veterinary medicine but are not lining up to become engineers?», *Cornell Chronicle*, 12 de junio de 2007, <https://news.cornell.edu/stories/2007/06/why-women-become-veterinarians-not-engineers>.

25. Bar-Hillel, «The base-rate fallacy in probability judgments», *Acta Psychologica*, 44, 1980, págs. 211-233, <https://doi.org/10.1016/00016918(80)90046-3>; Tversky y Kahneman, «Evidential impact...» art. cit.; Gigerenzer, «How to make...», art. cit.

26. Gigerenzer, «How to make...», art. cit.; «On narrow norms and vague heuristics: A reply to Kahneman and Tversky», *Psychological Review*, 103, 1996, págs. 592-596, <https://doi.org/10.1037/0033-295X.103.3.592>; Kahneman y Tversky, «On the reality...», art. cit.

27. Cosmides y Tooby, «Are humans...», art. cit.; Gigerenzer, «How to make...», art. cit.; Hoffrage, Lindsey, Hertwig y Gigerenzer, «Communicating statistical information», *Science*, 290, 2000, págs. 2261-2262, <https://doi.org/10.1126/science.290.5500.2261>; Tversky y Kahneman, «Extensions versus intuitive reasoning...», art. cit. Kahneman y Tversky señalan que los formatos de frecuencias reducen, pero no siempre eliminan, la falacia de la tasa base, como vimos en el capítulo 1 con la colaboración contradictoria de Kahneman con el colaborador de Gigerenzer, Ralph Hertwig; sobre si los formatos de frecuencias eliminan la falacia de la conjunción: Kahneman y Tversky, «On the reality...», art. cit.; Mellers, Hertwig y Kahneman, «Do frequency...», art. cit.

28. Gigerenzer, «On the supposed evidence...», art. cit.; Kahan, «Ideology, motivated...», art. cit.

1. El modelo del ser humano como actor racional se explica en cualquier manual introductorio de economía o ciencias políticas. La teoría que relaciona la elección racional con la utilidad esperada fue desarrollada por Von Neumann y Morgenstern, *Theory of Games and Economic Behavior*, Princeton (NJ), Princeton University Press, 2007 [1953], y refinada por Savage, *The Foundations...*, *op. cit*. Yo utilizaré indistintamente «elección racional» y «utilidad esperada» para la teoría que las equipara. Para unas explicaciones accesibles, véase Luce y Raiffa, *Games and Decisions: Introduction and Critical Survey*, Nueva York, Dover, 1957, y Hastie y Dawes, *Rational Choice...*, *op. cit*.

2. Cohn, Maréchal, Tannenbaum y Zünd, «Civic honesty around the globe», *Science*, 365, 2019, págs. 70-73, <https://doi.org/10.1126/science.aau8712>.

3. Glaeser, «Psychology and the market», *The American Economic Review*, 94, 2004, págs. 408-413, <http://www.jstor.org/stable/3592919>.

4. Sobre el cuestionamiento de los axiomas de la elección racional: Arkes, Gigerenzer y Hertwig, «How bad is incoherence?», *Decision*, 3, 2016, págs. 20-39, <https://doi.org/10.1037/dec0000043>; Slovic y Tversky, «Who accepts Savage's axiom?», *Behavioral Science*, 19, 1974, págs. 368-373, <https://doi.org/10.1002/bs.3830190603>.

5. Hastie y Dawes, *Rational Choice..., op. cit*.; Savage, *The Foundations..., op. cit*.

6. Más comúnmente se denomina *completitud* o *comparabilidad*.

7. Conocida asimismo como *distribución de probabilidades entre alternativas, álgebra de combinación* y *reducción de loterías compuestas*.

8. Entre las variantes del axioma de independencia se incluyen la condición de Chernoff, la propiedad de Sen, la independencia de las alternativas irrelevantes de Arrow (IAI) y el axioma de elección de Luce.

9. Liberman, «If P...», art. cit.

10. Más comúnmente, *continuidad* o *solvencia*.

11. Stevenson y Wolfers, «Economic growth and subjective well-being: Reassessing the Easterlin Paradox», *Brookings Papers on Economic Activity*, 1-87, 2008, <https://doi.org/10.3386/w14282>.

12. Richardson, *Statistics of Deadly Quarrels*, Pittsburgh, Boxwood Press, 1960, pág. 11; Slovic, «"If I look at the mass I will never act": Psychic numbing and genocide», *Judgment and Decision Making*, 2, 2007, págs. 79-95, <https://doi.org/10.1007/978-90-481-8647-1_3>; Wan y Shammas, «Why Americans are numb to be the staggering coronavirus death toll», *The Washington Post*, 21 de diciembre de 2020, <https://www.washingtonpost.com/health/2020/12/21/covid-why-we-ignoredeaths/>.

13. Pinker, *The Better Angels...*, *op. cit*., págs. 219-220.

14. Tetlock, «Thinking the unthinkable...», art. cit.; Tetlock, Kristel, Elson, Green y Lerner, «The psychology...», art. cit.

15. «¡Caramba, un millón de dólares...!, tal vez.» «¿Te acostarías conmigo por cien dólares?» «¿Qué clase de mujer te crees que soy?» «Eso ya lo hemos establecido; solo estamos regateando el precio.»

16. Simon, «Rational choice and the structure of the environment», *Psychological Review*, 63, 1956, págs. 129-138, <https://doi.org/10.1037/h0042769>.

17. Tversky, «Elimination by aspects: A theory of choice», *Psychological Review*, 79, 1972, págs. 281-299, <https://doi.org/10.1037/h0032955>.

18. Savage, *The Foundations...*, *op. cit*., citado en Tversky, «Elimination by aspects...», art. cit., págs. 283-284.

19. Tversky, «Intransitivity of preferences», *Psychological Review*, 76, 1969, págs. 31-48, <https://doi.org/10.1037/h0026750>.

20. Arkes, Gigerenzer y Hertwig, «How bad...», art. cit.

21. Tversky, «Elimination by aspects...», art. cit., pág. 298; Hastie y Dawes, *Rational Choice*..., *op. cit*., pág. 251.

22. Llamado *inversiones de preferencias*: Lichtenstein y Slovic, «Reversals of preference between bids and choices in gambling decisions», *Journal of Experimental Psychology*, 89, 1971, págs. 46-55, <https://doi.org/10.1037/h0031207>.

23. Redondeando los resultados en una diferencia de uno o dos centavos, pero las diferencias se anulan en las apuestas utilizadas en el estudio y no afectan a los resultados.

24. Bombas de dinero no intransitivas: Arkes, Gigerenzer y Hertwig, «How bad...», art. cit., pág. 23. Bombas de dinero de inversión de preferencias: Hastie y Dawes, *Rational Choice..., op. cit*., pág. 76. Abrir los ojos en Arkes, Gigerenzer y Hertwig, «How bad...», art. cit., págs. 23-24.

25. Allais, «Le comportement de l’homme rationnel devant le risque: Critique des postulats et axiomes de l’école americaine», *Econometrica*, 21, 1953, págs. 503-546, <https://doi.org/10.2307/1907921>.

26. Kahneman y Tversky, «Prospect theory: An analysis of decisions under risk», *Econometrica*, 47, 1979, págs. 313-327, <https://doi.org/10.1142/9789814417358_0006>, pág. 267.

27. *Ibidem*.

28. Breyer, *Breaking the Vicious Circle: Toward Effective Risk Regulation*, Cambridge (MA), Harvard University Press, 1993, pág. 12.

29. Kahneman y Tversky, «Prospect theory...», art. cit.

30. McNeil, Pauker, Sox y Tversky, «On the elicitation of preferences for alternative therapies», *New England Journal of Medicine*, 306, 1982, págs. 1259-1262, <https://doi.org/10.1056/NEJM198205273062103>.

31. Tversky y Kahneman, The framing of decisions and the psychology of choice», *Science*, 211, 1981, págs. 453-458, <https://doi.org/10.1126/science.7455683>.

32. Hastie y Dawes, *Rational Choice...*, *op. cit*., págs. 282-288.

33. Kahneman y Tversky, «Prospect theory...», art. cit.

34. La gráfica del peso de las decisiones difiere de la figura 4 en Kahneman y Tversky, «Prospect theory...», art. cit., y se basa en cambio en la figura 12.2 en Hastie y Dawes, *Rational Choice...*, *op. cit.*, que a mi juicio es una mejor visualización de la teoría.

35. Basado en Kahneman y Tversky, «Prospect theory...», art. cit.

36. Esta asimetría generalizada se denomina *sesgo de negatividad*; Tierney y Baumeister, *The Power of Bad: How the Negativity Effect Rules Us and How We Can Rule It*, Nueva York, Penguin, 2019.

37. Maurice Allais, Herbert Simon, Daniel Kahneman, Richard Thaler y George Akerlof.

38. Gigerenzer, *Rationality for Mortals: How People Cope with Uncertainty*, Nueva York, Oxford University Press, 2008, pág. 20.

39. Abito y Salant, «The effect of product misperception on economic outcomes: Evidence from the extended warranty market», *Review of Economic Studies*, 86, 2018, págs. 2285-2318, <https://doi.org/10.1093/restud/rdy045>; Braverman, «Why you should steer clear of extended warranties», *Consumer Reports*, 22 de diciembre de 2018, <https://www.consumerreports.org/extended-warranties/steer-clear-extended-warranties/>.

40. Sydnor, «(Over)insuring Modest Risks», *American Economic Journal: Applied Economics*, 2, 2010, págs. 177-199.

41. Gigerenzer y Kolpatzik, «How new fact boxes are explaining medical risk to millions», *BMJ*, 357, 2017, j2460, <https://10.1136/bmj.j2460>; para un argumento similar sobre la detección del cáncer de pecho, véase también Gigerenzer, «Breast cancer screening pamphlets mislead women», *BMJ*, 348, 2014, g2636, <https://doi.org/10.1136/bmj.g2636>.

1. Twain, *Following the Equator*, Nueva York, Dover, 1989 [1897] (trad. cast.: *Viaje alrededor del mundo, siguiendo el Ecuador*, Barcelona, Laertes, 1992).

2. Sobre la teoría de la detección de señales y la teoría de la utilidad esperada, Wormwood, Barrett y Quigley, «Decision making from economic and signal detection perspectives: Development of an integrated framework», *Frontiers in Psychology*, 6, 2015, <https://doi.org/10.3389/fpsyg.2015.00952>.

3. Las distribuciones estadísticas se explican en cualquier introducción a la estadística o a la psicología. La teoría de la detección de señales, en Green y Swets, *Signal Detection Theory and Psychophysics*, Nueva York, Wiley, 1966; Lynn, Wormwood, Barrett y Quigley, «Decision making...», art. cit.; Swets, Dawes y Monahan, «Better decisions through science», *Scientific American*, 283, 2000, págs. 82-87; Wolfe, Kluender, Levi *et al.*, *Sensation & Perception*, Sunderland (MA), Sinauer, 2020 (6.ª ed.), cap. 1. Para las historias de la teoría de la detección de señales y la teoría estadística de la decisión y la conexión entre ambas, véase Gigerenzer, Krauss y Vitouch, «The null ritual: What you always wanted to know about significance testing but were afraid to ask», en D. Kaplan (comp.), *The Sage Handbook of Quantitative Methodology for the Social Sciences*, Thousand Oaks (CA), Sage, 2004; Gigerenzer, Swijtink, Porter *et al.*, *The Empire of Chance...*, *op. cit*.

4. Pinker, *The Better Angels...*, *op. cit.*, págs. 210-220.

5. Se denomina *teorema del límite central*.

6. *Verosimilitud* se utiliza aquí en el sentido restringido habitual en las discusiones de la regla de Bayes.

7. Lynn, Wormwood, Barrett y Quigley, «Decision making...», art. cit.

8. *Ibidem*.

9. *Ibidem*.

10. De manera confusa, *sensibilidad* se utiliza en los contextos médicos para referirse a la tasa de aciertos, esto es, a la verosimilitud de un resultado positivo cuando está presente la enfermedad. Contrasta con la *especificidad*, la tasa de rechazos correctos, es decir, la probabilidad de un resultado negativo cuando está ausente la enfermedad.

11. Loftus, Doyle, Dysart *et al*., *Eyewitness Testimony: Civil and Criminal*, Dayton (OH), LexisNexis, 2019 (6.ª ed.).

12. National Research Council, *Strenghtening Forensic Science in the United States: A Path Forward*, Washington, D. C., National Academies Press, 2009; President's Council of Advisors on Science and Technology, *Report to the President: Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*, 2016, <https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf>.

13. Cuestionamiento del interrogatorio mejorado: Bankoff, «Dick Cheney simply does not care that the CIA tortured innocent people», *The New York Magazine*, 14 de diciembre de 2014, <https://nymag.com/intelligencer/2014/12/cheney-alright-with-torture-of-innocent-people.html>.

14. Ali, *Dear Colleague Letter* (orientaciones normativas del subsecretario de derechos civiles), US Department of Education, 2011, <https://www2ed.gov/about/offices/list/ocr/letters/colleague-201104.html>.

15. Cuestionamiento de la conducta sexual inapropiada, en Soave, «Ezra Klein “completely supports” “terrible” Yes Means Yes law», *Reason*, 13 de octubre de 2014, <https://reason.com/2014/10/13/ezra-klein-completelysupportsterrible/>; Young, «The argument against affirmative con sent laws gets Voxjacked», *Reason*, 15 de octubre de 2014, <https://reason.com/2014/10/15/theargumentagainstaffirmativeconsent/>. Dos estudios de falsas acusaciones de violación han revelado tasas de entre el 5 % y el 10 %: De Zutter, Horselenberg y Van Koppen, «The prevalence of false allegations of rape in the United States from 2006-2010», *Journal of Forensic Psychology*, 2, 2017, <https://doi.org/10.4172/2475-319X.1000119>; Rumney, «False allegations of rape», *The Cambridge Law Journal*, 65, 2006, págs. 128-158, <https://doi.org/10.1017/S0008197306007069>. Véase también Bazelon y Larimore, «How often do women falsely cry rape?», *Slate*, 1 de octubre de 2009, <https://slate.com/news-and-politics/2009/10/why-it-sso-hard-to-quantify-false-rape-charges.html>, y Young, Crying rape», *Slate*, 18 de septiembre de 2014, <https://slate.com/human-interest/2014/09/false-rape-accusations-why-must-we-pretend-they-never-happen.html>.

16. Arkes y Mellers, «Do juries meet our expectations?», *Law and Human Behavior*, 26, 2002, págs. 625-639, <https://doi.org/10.1023/A:1020929517312>.

17. Arkes y Mellers citan un estudio de 1981 que refería un rango de 0,6 a 0,9, y un conjunto de estudios defectuosos con las *d’* más próximas a 2,7. Mi estimación procede de un metaanálisis del National Research Council, *The Polygraph and Lie Detection*, Washington, D. C., National Academies Press, 2003, pág. 122, que refiere una mediana de 0,86 para una medida relacionada de la sensibilidad, área bajo la curva ROC. Esa cifra puede convertirse, bajo el supuesto de distribuciones normales de igual varianza, en una *d’* de 1,53 multiplicando la correspondiente puntuación *z* por $\sqrt{2}$.

18. Sobre las falsas acusaciones, condenas y ejecuciones: National Research Council, *Strenghtening Forensic..., op. cit.*; President's Council of Advisors on Science and Technology, *Report to the President..., op. cit*. Para la violación en particular: Bazelon y Larimore, «How often...», art. cit.; De Zutter, Horselenberg y Van Koppen, «The prevalence...», art. cit.; Rumney, «False allegations...», art. cit.; Young, «Crying rape», art. cit. Para el terrorismo, Mueller, *Overblown..., op. cit*.

19. La teoría estadística de la decisión, y en particular la prueba de significancia de la hipótesis nula, se explica en todos los manuales de estadística y psicología. Para su historia y su relación con la teoría de la detección de señales, véanse Gigerenzer, Krauss y Vitouch, «The null ritual...», art. cit.; Gigerenzer, Swijtink, Porter *et al*., *The Empire of Chance...*, *op. cit*.

20. Gigerenzer, Krauss y Vitouch, «The null ritual...», art. cit.

21. Al igual que en la nota 6, *verosimilitud* se utiliza en el sentido restringido habitual en las discusiones de la regla de Bayes, a saber, la probabilidad de los datos dada una hipótesis.

22. Gigerenzer, «Statistical rituals...», art. cit.; Open Science Collaboration, «Estimating the reproductibility...», art. cit.; Ioannidis, «Why most published...», art. cit.; Pashler y Wagenmakers, «Editors' introduction...», art. cit.

23. Véase <https://xkcd.com/882/>.

24. Nature (grupo editor), «In praise of replication studies and null results», *Nature*, 578, 2020, págs. 489-490, <https://doi.org/10.1038/d41586020-00530-6>. *«Nothing that is not there and the nothing that is»* está tomado de «The Snow Man» de Wallace Stevens.

25. Henderson, «The problem of induction», en E. N. Zalta (comp.), *The Stanford Encyclopedia of Philosophy*, 2020, <https://plato.stanford.edu/archives/spr2020/entries/induction-problem>; Hume, *An Enquiry Concerning...*, *op. cit*.

1. Hume, *A Treatise of Human Nature*, *op. cit.*, libro III, parte II, sección V, «Of the obligation of promises».

2. Von Neumann y Morgenstern, *Theory of Games...*, *op. cit*. Sobre las introducciones semitécnicas: Binmore, *Fun and Games: A Text on Game Theory*, Boston, Houghton Mifflin, 1991 (trad. cast.: *Teoría de juegos*, Madrid, McGraw-Hill, 1996); Luce y Raiffa, *Games and Decisions...*, *op. cit*. Sobre las no técnicas en su mayor parte: Binmore, *Game Theory: A Very Short Introduction*, Nueva York, Oxford University Press, 2007 (trad. cast.: *La teoría de juegos: una breve introducción*, Madrid, Alianza, 2011); Rosenthal, *The Complete Idiot's Guide to Game Theory*, Nueva York, Penguin, 2011. Sobre las completamente no técnicas: Poundstone, *Prisoner's Dilemma: John von Neumann, Game Theory, and the Puzzle of the Bomb*, Nueva York, Anchor, 1992 (trad. cast.: *El dilema del prisionero: John von Neumann*, *la teoría de juegos y la bomba*, Madrid, Alianza, 2015).

3. Cada juego presentado en este capítulo se examina en la mayoría de las fuentes de la nota 2.

4. Clegg, «Protean free will», manuscrito inédito, California Institute of Technology, 2012, <https://resolver.caltech.edu/CaltechAUTHORS:20120328-152031480>; Dennett, *Intuition Pumps and Other Tools for Thinking*, Nueva York, W. W. Norton, 2013 (trad. cast.: *Bombas de intuición y otras herramientas de pensamiento*, México, FCE, 2015), cap. 8.

5. Thomas, De Freitas, DeScioli y Pinker, «Recursive mentalizing and common knowledge in the bystander effect», *Journal of Experimental Psychology: General*, 145, 2016, págs. 621-629, <https://doi.org/10.1037/xge0000153>.

6. Chwe, *Rational Ritual...*, *op. cit*.; De Freitas, Thomas, DeScioli, Haque y Pinker, «The psychology of coordination and common knowledge», *Journal of Personality and Social Psychology*, 107, 2014, págs. 657-676, <https://doi.org/10.1037/a0037037>; Schelling, *The Strategy...*, *op. cit*.; Thomas, DeScioli y Pinker, «Common knowledge...», art. cit.

7. Pinker, *The Language...*, *op. cit.*, cap. 8; Schelling, *The Strategy...*, *op. cit.*

8. Lewis, *Convention: A Philosophical Study*, Cambridge (MA), Harvard University Press, 1969. Sobre el escepticismo acerca de que las convenciones requieran un conocimiento común, Binmore, «Do conventions need to be common knowledge?», *Topoi*, 27, 2008, págs. 17-27, <https://doi-org.ezp-prod1.hul.harvard.edu/10.1007/s11245-008-9033-4>.

9. Este ejemplo ha sido ajustado a la inflación.

10. Schelling, *The Strategy...*, *op. cit.*, págs. 67, 71.

11. Goldstein, «Chicken dilemmas...», art. cit.

12. Frank, *Passions...*, *op. cit.*; Schelling, *The Strategy...*, *op. cit.*; véase también Pinker, *How the Mind...*, *op. cit.*, cap. 6.

13. Subasta de dólares, en Poundstone, *Prisoner's Dilemma...*, *op. cit.*; Shubik, «The dollar auction game: A paradox in noncooperative behavior and escalation», *Journal of Conflict Resolution*, 15, 1971, págs. 109-111, <https://doi.org/10.1177/002200277101500111>.

14. Dawkins, *The Selfish Gene*, Nueva York, Oxford University Press, 2016 [1976] (trad. cast.: *El gen egoísta: las bases biológicas de nuestra conducta*, Barcelona, Salvat, 1994); Maynard Smith, *Evolution and the Theory of Games*, Nueva York, Cambridge University Press, 1982.

15. Pinker, *The Better Angels...*, *op. cit.*, págs. 217-220.

16. Shermer, «The doping dilemma: Game theory helps to explain the pervasive abuse of drugs in cycling, baseball, and other sports», *Scientific American*, abril de 2008, págs. 32-39.

17. Dawkins, *The Selfish Gene, op. cit*.; Maynard Smith, *Evolution..., op. cit*.

18. Trivers, «The evolution of reciprocal altruism», *The Quarterly Review of Biology*, 46, 1971, págs. 35-57, <https://doi.org/10.1086/406755>.

19. Pinker, *How the Mind...*, *op. cit*., cap. 7; Pinker, *The Blank Slate: The Modern Denial of Human Nature*, Nueva York, Penguin, 2016 [2002] (trad. cast.: *La tabla rasa: la negación moderna de la naturaleza humana*, Barcelona, Paidós, 2012), cap. 14; Pinker, *The Better Angels...*, *op. cit*., caps. 8 y 9; Trivers, «The evolution...», art. cit.

20. Ridley, *The Origins of Virtue: Human Instincts and the Evolution of Cooperation*, Nueva York, Viking, 1997.

21. Ellickson, *Order without Law: How Neighbors Settle Disputes*, Cambridge (MA), Harvard University Press, 1991; Ridley, *The Origins...*, *op. cit*.

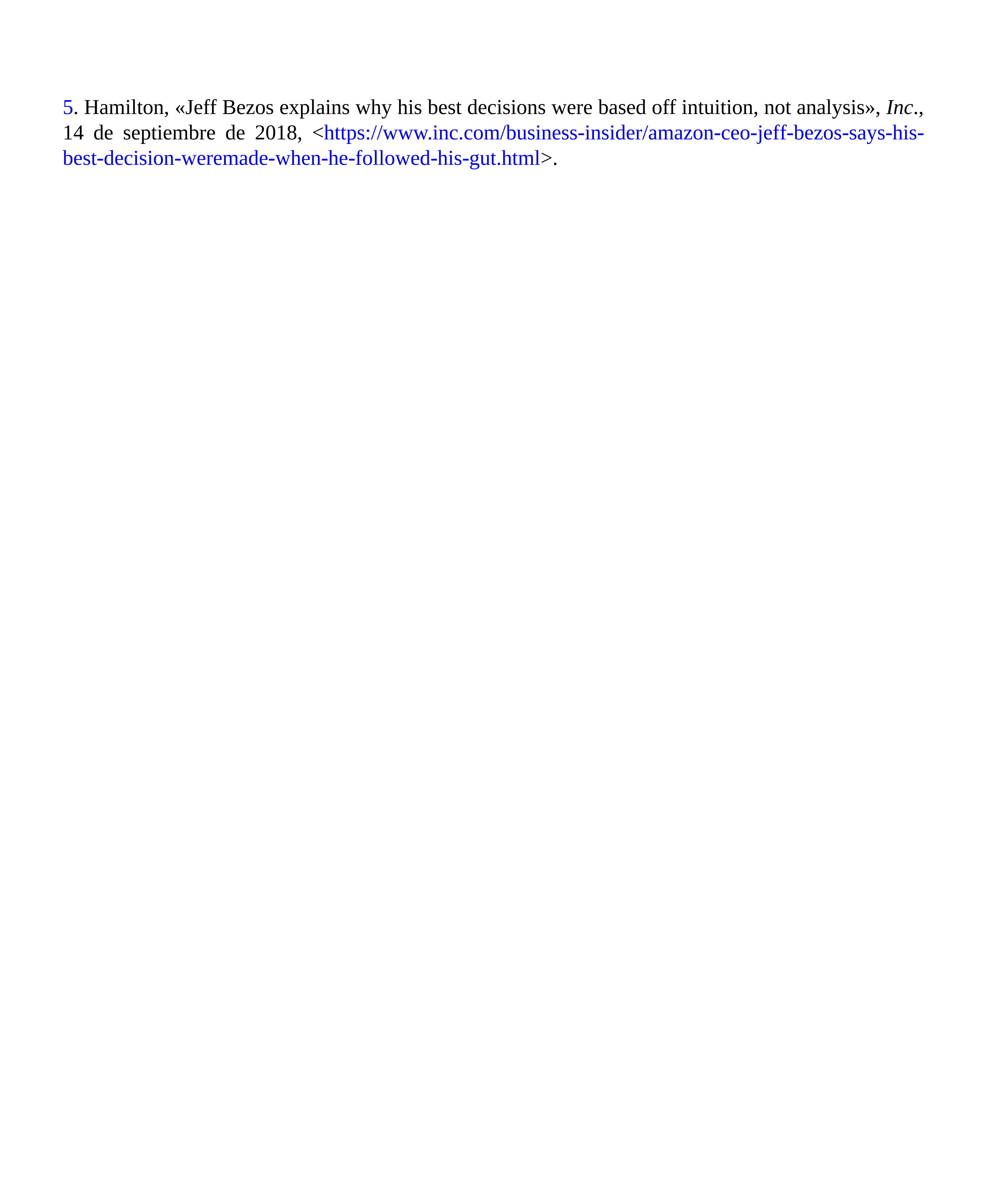
22. Hobbes, *Leviathan*, Nueva York, Oxford University Press, 1957 [1651] (trad. cast.: *Leviatán, o la materia, forma y poder de un Estado eclesiástico y civil*, Madrid, Alianza, 2014), cap. 14, pág. 190.

1. Sowell, *The Vision of the Anointed: Self-congratulation As a Basis for Social Policy*, Nueva York, Basic Books, 1995.

2. Cohen, *Science and the Founding Fathers: Science in the Political Thought of Thomas Jefferson, Benjamin Franklin, John Adams, and James Madison*, Nueva York, W. W. Norton, 1997.

3. BBC, «Avoid gold teeth, says Turkmen leader», *BBC News*, 7 de abril de 2004, <http://news.bbc.co.uk/2/hi/asiapacific/3607467.stm>.

4. Stevenson y Wolfers, «Economic growth...», art. cit., adaptado con permiso de los autores.



5. Hamilton, «Jeff Bezos explains why his best decisions were based off intuition, not analysis», *Inc.*, 14 de septiembre de 2018, <https://www.inc.com/business-insider/amazon-ceo-jeff-bezos-says-his-best-decision-weremade-when-he-followed-his-gut.html>.

6. Chapman y Chapman, «Genesis of popular but erroneous psychodiagnostic observations», *Journal of Abnormal Psychology*, 72, 1967, págs. 193-204, <https://doi.org/10.1037/h0024670>; «Illusory correlation as an obstacle to the use of valid psychodiagnostic signs», *Journal of Abnormal Psychology*, 74, 1969, págs. 271-280, <https://doi.org/10.1037/h0027592>.

7. Thompson y Adams, «The full moon and ED patient volumes: Unearthing a myth», *The American Journal of Emergency Medicine*, 14, 1996, págs. 161-164, <https://doi.org/10.1016/S0735-6757(96)90124-2>.

8. Sobre las correlaciones espurias, <https://www.tylervigen.com/spurious-correlations>.

9. Galton, «Regression toward mediocrity in hereditary stature», *Journal of the Anthropological Institute of Great Britain and Ireland*, 15, 1886, págs. 246-263.

10. Tversky y Kahneman, «Judgment under uncertainty...», art. cit.

11. *Ibidem*.

12. Tversky y Kahneman, «Belief in the law...», art. cit.; «Judgment under uncertainty...», art. cit.

13. El autor, Jonah Lehrer («The truth wears off», *The New Yorker*, 5 de diciembre de 2010, <https://www.newyorker.com/magazine/2010/12/13/thetruthwearsoff>) citaba a científicos que explicaban la regresión a la media y prácticas investigadoras que se le antojaban cuestionables, pero seguía manteniendo que algo estaba sucediendo, aunque no sabían de qué se trataba.

14. Pinker, *The Language...*, *op. cit.*, págs. 208-233.

15. Hume, *A Treatise of Human Nature*, *op. cit*.

16. Holland, Statistics and causal inference», *Journal of the American Statistical Association*, 81, 1986, págs. 945-960, <https://doi.org/10.2307/2289064>; King, Keohane y Verba, *Designing Social Inquiry: Scientific Inference in Qualitative Research*, Princeton (NJ), Princeton University Press, 1994 (trad. cast.: *El diseño de la investigación social: la inferencia científica en los estudios cualitativos*, Madrid, Alianza, 2018), cap. 3.

17. Kaba, «Yes, we mean literally abolish the police», *The New York Times*, 12 de junio de 2020, <https://www.nytimes.com/2020/06/12/opinion/sunday/floydabolishdefundpolice.html>. Para reseñas accesibles de estudios que sí muestran un efecto causal de la vigilancia policial sobre la delincuencia (empleando métodos explicados en este capítulo), véase Yglesias, «Defund police is a bad idea, not a bad slogan», *Slow Boring*, 7 de diciembre de 2020, <https://www.slowboring.com/p/defund-police-is-a-bad-idea-nota>, «The End of Policing left me convinced we still need policing», *Vox*, 18 de junio de 2020, <https://www.vox.com/2020/6/18/21293784/alex-vitaleend-of-policing-review>.

18. Pearl, *Causality: Models, Reasoning, and Inference*, Nueva York, Cambridge University Press, 2000.

19. Weissman, «Do GRE scores help predict getting a physics Ph.D.? A comment on a paper by Miller *et al*.», *Science Advances*, 6, 2020, eaax3787, <https://doi.org/10.1126/sciadv.aax3787>.

20. VanderWeele, «Commentary: Resolutions of the birthweight paradox: competing explanations and analytical insights», *International Journal of Epidemiology*, 43, 2014, págs. 1368-1373, <https://doi.org/10.1093/ije/dyu162>.

21. Letra de la grabación de 1941. Esto es lo que dice la Biblia, en Mateo 25:29: «Porque a quien tiene se le dará y le sobrará; pero a quien no tiene, aun lo que tiene se le quitará».

22. Social Progress Imperative, «2020 Social Progress Index», 2020, <https://www.socialprogress.org/>; Welzel, *Freedom Rising: Human Empowerment and the Quest for Emancipation*, Nueva York, Cambridge University Press, 2013.

23. Deary, *Intelligence: A Very Short Introduction*, Nueva York, Oxford University Press, 2001 (trad. cast.: *Una brevísima introducción a la inteligencia*, México, Océano, 2004); Temple, «The possible importance of income and education as covariates in cohort studies that investigate the relationship between diet and disease», *F1000Research*, 4, 2015, pág. 690, <https://doi.org/10.12688/f1000research.6929.2>; Ritchie, *Intelligence: All That Matters*, Londres, Hodders & Stoughton, 2015.

24. Pearl y Mackenzie, *The Book of Why...*, *op. cit*.

25. El psicólogo cognitivo Reid Hastie.

26. Baron, «Applying evidence to social programs», *The New York Times*, 29 de noviembre de 2012, <https://economix.blogs.nytimes.com/2012/11/29/applying-evidence-to-social-programs/>; Bornstein, «The dawn of the evidence-based budget», *The New York Times*, 30 de mayo de 2012, <https://opinionator.blogs.nytimes.com/2012/05/30/worthy-of-government-funding-prove-it>; Hallsworth y Kirkman, *Behavioral Insights..., op. cit*.

27. Levitt y Dubner, *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*, Nueva York, William Morrow, 2009 (trad. cast.: *Freakonomics: Un economista políticamente incorrecto explora el lado oculto de lo que nos afecta*, Barcelona, B de Bolsillo, 2018; <https://freakonomics.com/>.

28. DellaVigna y Kaplan, «The Fox News effect: Media bias and voting», *Quarterly Journal of Economics*, 122, 2007, págs. 1187-1234, <https://doi.org/10.1162/qjec.122.3.1187>.

29. Martin y Yurukoglu, «Bias in cable news: Persuasion and polarization», *American Economic Review*, 107, 2017, págs. 2565-2599, <https://doi.org/10.1257/aer.20160812>.

30. Véase Pinker, *The Better Angels...*, *op. cit*., págs. 278-284.

31. Este ejemplo es una adaptación de Russett y Oneal, *Triangulating Peace: Democracy, Interdependence, and International Organizations*, Nueva York, Norton, 2001, y se analiza en Pinker, *The Better Angels...*, *op. cit*., págs. 278-284.

32. Stuart, «Matching methods for causal inference: A review and a look forward», *Statistical Science*, 25, 2010, págs. 1-21, <https://doi.org/10.1214/09-STS313>.

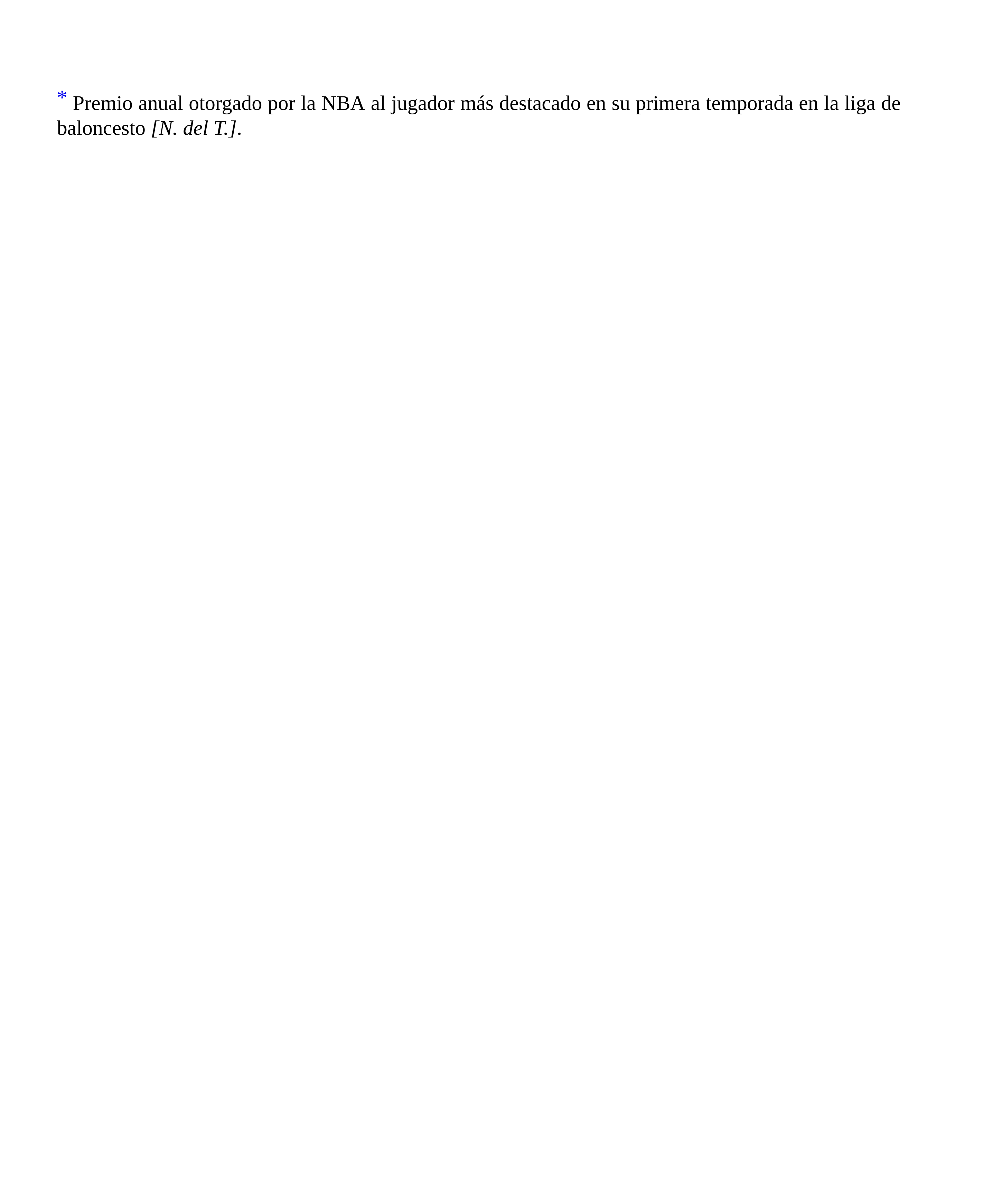
33. Kendler, Kessler, Walters *et al*., «Stressful life events, genetic liability, and onset of an episode of major depression in women», *Focus*, 8, 2010, págs. 459-470, <https://doi.org/10.1176/foc.8.3.foc.459>.

34. Vaci, Edelsbrunner, Stern *et al*., «The joint influence of intelligence and practice on skill development throughout the life span», *Proceedings of the National Academy of Sciences*, 116, 2019, págs. 18363-18369, <https://doi.org/10.1073/pnas.1819086116>.

35. Dawes, Faust y Meehl, «Clinical versus actuarial judgment», *Science*, 243, 1989, págs. 1668-1674, <https://doi.org/10.1126/science.2648573>; Meehl, *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*, Brattleboro, Vermont, Echo Point Books, 2013 [1954]. Véase también Tetlock, *Expert Political Judgment: How Good Is It? How Can We Know?*, Princeton (NJ), Princeton University Press, 2009, a propósito de las predicciones políticas y económicas.

36. Polderman, Benyamin, Leeuw *et al*., «Meta-analysis of the heritability of human traits base on fifty years of twin studies», *Nature Genetics*, 47, 2015, págs. 702-709, <https://doi.org/10.1038/ng.3285>; véase Pinker, *The Modern Denial..., op. cit*., págs. 395-398, 450-451.

37. Salganik, Lundberg, Kindel *et al*., «Measuring the predictability of life outcomes with a scientific mass collaboration», *Proceedings of the National Academy of Sciences*, 117, 2020, págs. 8398-8403, <https://doi.org/10.1073/pnas.1915006117>.



* Premio anual otorgado por la NBA al jugador más destacado en su primera temporada en la liga de baloncesto *[N. del T.]*.



* Nombre del personaje interpretado por Ali MacGraw en la película *Love Story* (Arthur Hiller, 1970) *[N. del T.]*.



* *«Them that's got shall get, them that's not shall lose. So the Bible said, and it still is news» [N. del T.].*

1. Shermer, «COVID-19 conspiracists and their discontents», *Quillette*, 7 de mayo de 2020, <https://quillette.com/2020/05/07/covid-19-conspiracists-and-their-discontents/>.

2. O’Keefe, «One in three Americans would not get COVID-19 vaccine», *Gallup*, 7 de agosto de 2020, <https://gallup.com/poll/317018/onethree-americans-not-covid-vaccine.aspx>.

3. Wolfe y Dale, «“It’s going to disappear”: A timeline of Trump’s claims that Covid-19 will vanish», CNN, 31 de octubre de 2020, <https://www.cnn.com/interactive/2020/10/politics/covid-disappearing-trump-comment-tracker/>.

4. Kessler, Rizzo y Kelly, «Trump is averaging more than 50 false or misleading claims a day», *The Washington Post*, 22 de octubre de 2020, <https://www.washingtonpost.com/politics/2020/10/22/presidenttrump-is-averaging-more-than-50-false-or-misleading-claims-day/>; Nature (grupo editor), «A four-year timeline of Trump's impact on science», *Nature*, 5 de octubre de 2020, <https://doi.org/10.1038/d41586-020-02814-3>; Tollefson, «How Trump damaged science – And why it could take decades to recover», *Nature*, 586, 5 de octubre de 2020, págs. 190-194, <https://www.nature.com/articles/d41586-020-02800-9>.

5. Rauch, *The Constitution of Knowledge: A Defense of Truth*, Washington, D. C., Brookings Institution Press, 2021.

6. Gilbert, «The 10 most-viewed fake-news stories on Facebook in 2019 were just revealed in a new report», *Business Insider*, 6 de noviembre de 2019, <https://www.businessinsider.com/most-viewed-fake-news-storiesshared-on-facebook-2019-2019-11>; Pennycook y Rand, «The cognitive science of fake news», Psyarxiv, 2020, <https://psyarxiv.com/ar96c>.

7. Las cinco primeras cifras son de una encuesta de Gallup (Moore, «Three in four Americans...», art. cit.); las cinco segundas son del Pew Forum on Religion and Public Life, *Many Americans...*, *op. cit*.

8. Según reiteradas encuestas, entre 1990 y 2005 o 2009 hubo ligeras tendencias ascendentes en la creencia en la sanación espiritual, las casas encantadas, los fantasmas, la comunicación con los muertos y las brujas, y tendencias ligeramente descendentes en la creencia en la posesión por el demonio, la percepción extrasensorial, la telepatía y la reencarnación. Las consultas a adivinos o videntes, la creencia en alienígenas que visitan la Tierra y la canalización se mantuvieron estables (Moore, «Three in four Americans...», art. cit.; Pew Forum on Religion and Public Life *Many Americans...*, *op. cit*.). Según los informes de la National Science Foundation, desde 1979 hasta 2018, el porcentaje de creyentes en que la astrología es «muy» o «más o menos» científica descendió muy levemente, desde el cuarenta y pocos hasta el treinta y muchos por ciento, y en 2018 incluía el 58 % de las personas entre los dieciocho y los veinticuatro años y el 49 % de los individuos de veinticinco a treinta y cuatro años (National Science Board, *Science and Engineering Indicators 2014*, Alexandria (VI) National Science Foundation, 2014, <https://www.nsf.gov/statistics/seind14/index.cfm/home>; *The State of U.S. Science and Engineering 2020*, Alexandria (VI), National Science Foundation, 2020, <https://ncses.nsf.gov/pubs/nsb20201/>). Todas las creencias paranormales son más populares entre los encuestados más jóvenes que entre los mayores (Pew Forum on Religion and Public Life, *Many Americans...*, *op. cit*.). En lo que concierne a la astrología, el gradiente de edad se mantiene estable a lo largo de las décadas, lo cual sugiere que la credulidad es un efecto de la propia juventud, y que muchas personas la pierden, y no depende de pertenecer a la generación Z, a los *millennials* o a cualquier otra cohorte.

9. Shermer, *Why People Believe Weird Things*, Nueva York, Freeman, 1997 (trad. cast.: *Por qué creemos en cosas raras: pseudociencia, confusión y otras supersticiones de nuestro tiempo*, Barcelona, Alba, 2008); *The Believing Brain: From Ghosts and Gods to Politics and Conspiracies*, Nueva York, St. Martin's Press, 2011; «The top ten weirdest things coutndown», *Skeptic*, 2020, <https://www.skeptic.com/reading_room/the-top-10-weirdestthings/>.

10. Mercier, *Not Born Yesterday...*, *op. cit.*; Shermer, «Why people believe conspiracy theories», *Skeptic*, 25, 2020, págs. 12-17; Sunstein y Vermeule, «Conspiracy theories», *John M. Olin Program in law and Economics Working Papers*, 387, 2008, <https://dx.doi.org/10.2139/ssrn.1084585>; Uscinski y Parent, *American Conspiracy Theories*, Nueva York, Oxford University Press, 2014; Van Prooijen y Van Vugt, «Conspiracy theories: Evolved functions and psychological mechanisms», *Perspectives on Psychological Science*, 13, 2018, págs- 770-788, <https://doi.org/10.1177/1745691618774270>.

11. Horowitz, *The Deadly Ethnic Riot*, Berkeley, University of California Press, 2001; Sunstein y Vermeule, «Conspiracy...», art. cit.

12. Statista Research Department, «Beliefs and conspiracy theories in the U.S. – Statistics & Facts», Statista, 13 de agosto de 2019, <https://www.statista.com/topics/5103/beliefs-and-superstition-in-the-us/#dossierSummary_chapter5>; Uscinski y Parent, *American Conspiracy...*, *op. cit*.

13. Brunvand, *Too Good to Be True: The Colossal Book of Urban Legends*, Nueva York, W. W. Norton, 2014 (trad. cast.: *El fabuloso libro de las leyendas urbanas*, Barcelona, Debolsillo, 2008); los titulares de tabloides son de mi colección personal.

14. Nyhan, «Fake news and bots may be worrisome, but their political power is overblown», *The New York Times*, 13 de febrero de 2018, <https://www.nytimes.com/2008/02/13/upshot/fake-news-and-bots-may-be-worrisome-but-their-political-power-is-overblown.html>.

15. Goldstein, *36 Arguments for the Existence of God: A Work of Fiction*, Nueva York, Pantheon, 2010.

16. Véase <https://quoteinvestigator.com/2017/11/30/salary/>.

17. Kunda, «The case for motivated reasoning», *Psychological Bulletin*, 108, 1990, págs. 480-498, <https://doi.org/10.1037/0033-2909.108.3.480>.

18. Gracias a la lingüista Ann Farmer por su lema «No se trata de tener razón. Se trata de hacer bien las cosas».

19. Aunque véase la nota 26 del cap. 1.

20. Dawson, Gilovich y Regan, «Motivated reasoning...», art. cit.

21. Kahan, Peters, Dawson y Slovic, «Motivated numeracy and enlightened self-goverment», *Behavioural Public Policy*, 1, 2017, págs. 54-86, <10.1017/bpp.2016.2>; Lord, Ross y Lepper, «Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence», *Journal of Personality and Social Psychology*, 37, 1979, págs. 2098-2109, <https://doi.org/10.1037/0022-3514.37.11.2098>; Taber y Lodge, «Motivated skepticism in the evaluation of political beliefs», *American Journal of Political Science*, 50, 2006, págs. 755-769, <https://doi.org/10.1111/j.1540-5907.2006.00214.x>, Dawson, Gilovich y Regan, «Motivated reasoning...», art. cit.

22. Pronin, Lin y Ross, «The bias blind spot: Perceptions of bias in self versus others», *Personality and Social Psychology Bulletin*, 28, 2002, págs. 369-381, <https://doi.org/10.1177/0146167202286008>.

23. Mercier y Sperber, «Why do humans...», art. cit., 2017; Tetlock, «Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors», *Psychological Review*, 109, 2002, págs. 451-471, <https://doi.org/10.1037/0033-295X.109.3.451>. Pero véase también Norman, «Why we reason...», art. cit.

24. Mercier y Sperber, «Why do humans...», art. cit., pág. 63; Mercier, Trouche, Yama, *et al*., «Experts and laymen grossly underestimate the benefits of argumentation for reasoning», *Thinking & Reasoning*, 21, 2015, págs. 341-355, <https://doi.org/10.1080/13546783.2014.981582>.

25. Kahan, Peters, Dawson y Slovic, «Motivated numeracy...», art. cit.

26. Ditto, Liu, Clark *et al*., «At least bias is bipartisan: A metaanalytic comparison of partisan bias in liberals and conservatives», *Perspectives on Psychological Science*, 14, 2019, págs. 273-291, <https://doi.org/10.1177/1745691617746796>. Para las respuestas, véase Baron y Jost, «False equivalence: Are liberals and conservatives in the United States equally biased?», *Perspectives on Psychological Science*, 14, 2019, págs. 292-303, <https://doi.org/10.1177/1745691618788876>; Ditto, Clark, Liu *et al*., «Partisan bias and its discontents», *Perspectives on Psychological Science*, 14, 2019, págs. 304-316, <https://doi.org/10.1177/1745691618817753>.

27. Stanovich, «The bias that divides us», *Quillette*, 26 de septiembre de 2020, <https://quillette.com/2020/09/26/the-bias-that-devides-us/>; *The Bias That Divides Us: The Science and Politics of Myside Thinking*, Cambridge (MA), MIT Press, 2021.

28. Gampa, Wokcik, Motyl *et al.*, «(Ideo)logical reasoning: Ideology impairs sound reasoning», *Social Psychological and Personality Science*, 10, 2019, págs.1075-1083, <https://doi.org/10.1177/1948550619829059>.

29. Kahan, Hoffman, Braman *et al*., «“They saw a protest”: Cognitive illiberalism and the speech-conduct distinction», *Stanford Law Review*, 64, 2012, págs. 851-906.

30. Kahan, Peters, Wittlin *et al.*, «The polarizing impact of science litteracy and numeracy on perceived climate change risks», *Nature Climate Change*, 2, 2012, págs. 732-735, <https://doi.org/10.1038/nclimate1547>.

31. Stanovich, «The bias...», art. cit., *The Bias That Divides...*, *op. cit*.

32. Jerárquico frente a igualitario y libertario frente a comunitarista, en Kahan, «Ideology, motivated...», art. cit., y otras referencias en la nota 39. Trono y altar frente a Ilustración, tribal frente a cosmopolita, en Pinker, *Enlightenment Now...*, *op. cit*., caps. 21, 23. Trágico frente a utópico, en Pinker, *The Blank Slate...*, *op. cit*.; Sowell, *A Conflict of Visions: Ideological Origins of Political Struggles*, Nueva York, Quill, 1987, cap. 16. Honor frente a dignidad, en Pinker, *The Better Angels...*, *op. cit*., cap. 3, Campbell y Manning, *The Rise of Victimhood Culture: Microaggressions, Safe Spaces, and the New Culture Wars*, Londres, Palgrave Macmillan, 2018; Pinker, «Why are states so red and blue?», *The New York Times*, 24 de octubre de 2012, <http://opinionator.blogs.nytimes.com/2012/10/24/why-are-states-so-red-and-blue/?_r=0>.
Vinculante frente a individualizador, en Haidt, *The Righteous Mind...*, *op. cit*.

33. Finkel, Bail, Cikara *et al*., «Political sectarianism in America», *Science*, 370, 2020, págs. 533-536, <https://doi.org/10.1126/science.abe1715>.

34. *Ibidem*; Wilkinson, *The Density Divide: Urbanization, Polarization, and Populist Backlash*, Washington, D. C., Niskanen Center, 2019, <https://www.niskanencenter.org/the-density-divide-urbanization-polarizationand-populist-backlash/>.

35. Baron y Jost, «False equivalence...», art. cit.

36. El epígrafe de Sowell, *The Vision...*, *op. cit*.

37. Ditto, Clark, Liu *et al*., «Partisan bias...», art. cit. Equivocaciones por ambas partes, en Pinker, *Enlightenment Now...*, *op. cit*., págs. 363-366.

38. Mercier, *Not Born Yesterday..., op. cit.*, págs. 191-197.

39. Kahan, «Ideology, motivated...», art. cit.; Kahan, Peters, Dawson *et al*., «Motivated numeracy...», art. cit; Kahan, Wittlin, Peters *et al*., «The tragedy of the riskperception commons: Culture conflict, rationality conflict, and climate change», *Yale Law & Economics Research Paper*, 435, 2011, <http://dx.doi.org/10.2139/ssrn.1871503>.

40. Mercier, *Not Born Yesterday...*, *op. cit*., cap. 10. Mercier citó la reseña de Google como ponente invitado en mi curso sobre racionalidad, 5 de marzo de 2020.

41. *Ibidem*; Sperber, «Intuitive and reflective beliefs», *Mind & Language*, 12, 1997, págs. 67-83, <https://doi.org/10.1111/j.1468-0017.1997.tb00062.x>.

42. Abelson, «Beliefs are like possessions», *Journal for the Theory of Social Behaviour*, 16, 1986, págs. 223-250, <https://doi.org/10.1111/j.1468-5914.1986.tb00078.x>.

43. Henrich, Heine y Norenzayan, «The weirdest people in the world?», *Behavioral and Brain Sciences*, 33, 2010, págs. 61-83, <https://doi.org/10.1017/S0140525X0999152X>.

44. Coyne, *Faith versus Fact: Why Science and Religion Are Incompatible*, Nueva York, Penguin, 2015; Dawkins, *The God Delusion*, Nueva York, Houghton Mifflin, 2006 (trad. cast.: *El espejismo de Dios*, Barcelona, Espasa, 2017); Dennett, *Breaking the Spell: Religion As a Natural Phenomenon*, Nueva York, Penguin, 2006 (trad. cast.: *Romper el hechizo: la religión como un fenómeno natural*, Madrid, Katz, 2011); Harris, *The End of Faith: Religion, Terror, and the Future of Reason*, Nueva York, W. W. Norton, 2005 (trad. cast.: *El fin de la fe: la religión, el terror y el futuro de la razón*, Madrid, Paradigma, 2007). Véase R. Goldstein, *36 Arguments...*, *op. cit*., para un debate novelado.

45. Jenkins, «The Crown’s fake history is as corrosive as fake news», *The Guardian*, 16 de noviembre de 2020, <http://www.theguardian.com/commentisfree/2020/nov/16/the-crown-fake-history-news-tv-series-royalfamily-artistic-licence>.

46. BBC, «The Crown: Netflix has “no plans” for a fiction warning», *BBC News*, 6 de diciembre de 2020, <https://www.bbc.com/news/entertainment-arts-55207871>.

47. Baumard y Boyer, «Religious beliefs as reflective elaborations on intuitions: A modified dual-process model», *Current Directions in Psychological Science*, 22, 2013, págs. 295-300, <https://doi.org/10.1177/0963721413478610>; Hood, *Supersense: Why we Believe in the Unbelievable*, Nueva York, HarperCollins, 2009; Pinker, *How the Mind...*, *op. cit*., caps. 5 y 8; Shermer, *Why People...*, *op. cit.*, *The Believing Brain:...*, *op. cit*.

48. Bloom, *Descartes' Baby: How the Science of Child Development Explains What Makes Us Human*, Nueva York, Basic Books, 2004.

49. Gelman, *The Essential Child: Origins of Essentialism in Everyday Thought*, Nueva York, Oxford University Press, 2005; Hood, *Supersense..., op. cit*.

50. Kelemen y Rosset, «The human function compunction: teleological explanation in adults», *Cognition*, 111, 2009, págs. 138-143, <https://doi.org/10.1016/j.cognition.2009.01.001>.

51. Rauch, *The Constitution...*, *op. cit*.; Shtulman, *Scienceblind: Why our Intuitive Theories about the World Are So Often Wrong*, Nueva York, Basic Books, 2017; Sloman y Fernbach, *The Knowledge Illusion: Why We Never Think Alone*, Nueva York, Penguin, 2017.

52. Véanse las revistas *Skeptical Inquirer* (<http://www.csicop.org/si>) y *Skeptic* (<http://www.skeptic.com/>), y el Center for Inquiry (<https://centerforinquiry.org/>) para las actualizaciones periódicas sobre pseudociencia en los medios de comunicación convencionales.

53. Acerbi, «Cognitive attraction and online misinformation», *Palgrave Communications*, 5, 2019, págs. 1-7, <https://doi.org/10.1057/s41599-0190224-y>.

54. Thompson, «QAnon is like a game – A most dangerous game», *WIRED Magazine*, 22 de septiembre de 2020, <https://www.wired.com/story/qanon-most-dangerous-multiplatform-game/>.

55. Mercier, *Not Born Yesterday...*, *op. cit*.; Shermer, «Why people believe...», art. cit.; Van Prooijen y Van Vugt, «Conspiracy theories...», art. cit.

56. Pinker, *The Better Angels...*, *op. cit.*, cap. 2; Chagnon, *Yanomamö...*, *op. cit.*

57. Van Prooijen y Van Vugt, «Conspiracy theories...», art. cit.

58. Mercier, *Not Born Yesterday..., op. cit.*, cap. 10.

59. Dawkins, *The Selfish Gene...*, *op. cit*.

60. Friesen, Campbell y Kay, «The psychological advantage of unfalsibiability: The appeal of untestable religious and political ideologies», *Journal of Personality and Social Psychology*, 108, 2015, págs. 515-529, <https://doi.org/10.1037/pspp0000018>.

61. Moore, «Three in four...», art. cit.; Pew Forum on Religion and Public Life, *Many Americans...*, *op. cit*.

62. Kahan, Climate-science communication and the measurement problem», *Political Psychology*, 36, 2015, págs. 1-43, <https://doi.org/10.1111/pops.12244>; Kahan, Wittlin, Peters *et al*., «The tragedy...», art. cit.

63. Nyhan y Reifler, «The roles of information deficits and identity threat in the prevalence of misperceptions», *Journal of Elections, Public Opinion and Parties*, 29, 2019, págs. 222-244, <https://doi.org/10.1080/17457289.2018.1465061>; Pennycook y Rand, «The cognitive science...», art. cit.; Wood y Porter, «The elusive backfire effect: Mass attitudes' steadfast factual adherence», *Political Behavior*, 41, 2019, págs. 135-163, <https://doi.org/10.1007/s11109-018-9443-y>.

64. Baron, «Actively open-minded thinking in politics», *Cognition*, 188, 2019, págs. 8-18, <https://doi.org/10.1016/j.cognition.2018.10.004>; Pennycook, Cheyne, Koehler y Fugelsang, «On the belief that beliefs should change according to evidence: Implications for conspirational, moral, paranormal, political, religious, and science beliefs», *Judgment and Decision Making*, 15, 2020, págs. 476-498, <https://doi.org/10.31234/osf.io/a7k96>; Sá, West y Stanovich, «The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill», *Journal of Educational Psychology*, 91, 1999, págs. 497-510, <https://doi.org/10.1037/0022-0663.91.3.497>; Tetlock y Gardner, *Superforecasting...*, *op. cit*.

65. Como la mayoría de las citas sucintas, de manera apócrifa; probablemente debería atribuírsele al colega economista Paul Samuelson: <https://quoteinvestigator.com/2011/07/22/keynes-change-mind/>.

66. Pennycook, Cheyne, Koehler y Fugelsang, «On the belief...», art. cit. Los tres primeros ítems fueron añadidos al test de apertura mental activa por Sá, West y Stanovich, «The domain...», art. cit.

67. Pennycook, Cheyne, Koehler y Fugelsang, «On the belief...», art. cit. Para hallazgos similares, véase Erceg, Galić y Bubić, «“Dysrationalia” among university students: The role of cognitive abilities, different aspects of rational thought and self-control in explaining epistemically suspect beliefs», *Europe’s Journal of Psychology*, 15, 2019, págs. 159-175, <https://doi.org/10.5964/ejop.v15i1.1696>; Stanovich, «On the distinction between rationality and intelligence: Implications for understanding individual differences in reasoning», en K. J. Holyoak y R. G. Morrison (comps.), *The Oxford Handbook of Thinking and Reasoning*, Nueva York, Oxford University Press, 2012. Pennycook, Cheyne, Seli *et al*., «Analytic cognitive style predicts religious and paranormal belief», *Cognition*, 123, 2012, págs. 335-346, <https://doi.org/10.1016/j.cognition.2012.03.003>; Stanovich, West y Toplak, *The Rationality Quotient: Toward a Test of Rational Thinking*, Cambridge (MA), MIT Press, 2016, y Stanovitch y Toplak, «The need for intellectual diversity in psychological science: Our own studies of actively open-minded thinking as a case study», *Cognition*, 187, págs. 156-166, 2019, <https://doi.org/10.1016/j.cognition.2019.03.006>, observan que algunas de estas correlaciones pueden ser exageradas por el término *creencia* en el cuestionario sobre la apertura, que algunos encuestados interpretan como *creencia religiosa*. Cuando se emplea la palabra *opinión*, las correlaciones son más bajas, pero todavía significativas.

68. Tendencias globales en creencias políticas y sociales, en Welzel, *Freedom Rising...*, *op. cit.*, Pinker, *Enlightenment Now...*, *op. cit.*, cap. 15.

69. Pennycook, Cheyne, Seli *et al*., «Analytic cognitive...», art. cit.; Stanovich, «On the distinction.», art. cit.; Stanovich, West y Toplak, *The Rationality Quotient...*, *op. cit*. Sobre el test de reflexión cognitiva: Frederick, «Cognitive reflection...», art. cit. Véase también Maymin y Langer, «Cognitive biases and mindfulness», *Humanities and Social Sciences Communications*, 8, pág. 40, 2021, <https://doi.org/10.1057/s41599-021-00712-1>, quienes lo conectan con la atención plena *(mindfulness)*.

70. Pennycook, Cheyne, Seli *et al.*, «Analytic cognitive...», art. cit. Pennycook y Rand, «Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking», *Journal of Personality*, 88, 2020, <https://doi.org/10.1111/jopy.12476>.

71. Sobre el sistema immune cognitivo, Norman, *Mental Immunity: Infectious Ideas, Mind Parasites, and the Search for a Better Way to Think*, Nueva York, HarperCollins, 2021.

72. Caplan, «What’s wrong...», art. cit.; *The AI Does...*, *op. cit*.; Raemon, «What exactly is the “Rationality Community”?», *LessWrong*, 9 de abril de 2017, <https://www.lesswrong.com/posts/s8yvtCbbZW2S4WnhE/whatexactly-is-the-rationality-community>.

73. La expresión *partido de estúpidos* se ha atribuido al exgobernador republicano de Luisiana Bobby Jindal, aunque lo que él dijo fue *partido estúpido*. Críticas desde dentro del movimiento conservador antes de Trump: M. K. Lewis, *Too Dumb to Fail: How the GOP Betrayed the Reagan Revolution to Win Elections (and How It Can Reclaim Its Conservative Roots)*, Nueva York, Hachette, 2016; Mann y Ornstein, *It's Even Worse Than It Looks: How the American Constitutional System Collided with the New Politics of Extremism*, Nueva York, Basic Books, 2016 [2012]; Sykes, *How the Right Lost Its Mind*, Nueva York, St. Martin's Press, 2017. Después de Trump: Saldin y Teles, *Never Trump: The Revolt of the Conservative Elites*, Nueva York, Oxford University Press, 2020; véase también The Lincoln Project, <https://lincolnproject.us/>.

74. Citado en Rauch, «The constitution of knowledge», *National Affairs*, otoño de 2018, <https://www.nationalaffairs.com/publications/detail/the-constitution-of-knowledge>.

75. Mercier, *Not Born Yesterday..., op. cit*.

76. Lane, «A truth reckoning: Why we’re holding those who lied for Trump accountable», *Forbes*, 7 de enero de 2021, <https://www.forbes.com/sites/randalllane/2021/01/07/a-truth-reckoning-why-were-holding-those-wholied-for-trump-accountable/?sh=5fedd2605710>.

77. Rauch, «The constitution...», art. cit., *The Constitution of Knowledge..., op. cit.*; Sloman y Fernbach, *The Knowledge Illusion..., op. cit*.

78. Sobre la estabilidad de la confianza en la ciencia: American Acade my of Arts and Sciences, *Perceptions of Science in America*, Cambridge (MA), American Academy of Arts and Sciences, 2018. Sobre el hundimiento de la confianza en la ciencia, Jones, «Confidence in higher education down since 2015», *Gallup Blog*, 9 de octubre de 2018, <https://news.gallup.com/opinion/gallup/242441/confidencehighereducationdown2015.aspx>.

79. Flaherty, «Failure to communicate: Professor suspended for saying a Chinese word that sounds like a racial slur in English», *Inside Higher Ed.*, 2020, <https://www.insidehighered.com/news/2020/09/08/professor-suspended-saying-chinese-word-sounds-english-slur>. Para otros ejemplos, véase Kors y Siverglate, *The Shadow University: The Betrayal of Liberty on America's Campuses*, Nueva York, Free Press, 1998; Lukianoff, *Unlearning liberty: Campus Censorship and the End of American Debate*, Nueva York, Encounter Books, 2012; Lukianoff y Haidt, *The Coddling of the American Mind: How Good Intentions and Bad Ideas Are Setting Up a Generation for Failure*, Nueva York, Penguin, 2018; y la Heterodox Academy, <https://heterodoxacademy.org/>, la Foundation for Individual Rights in Education, <https://www.thefire.org/>, y la revista *Quillette*, <https://quillette.com/>.

80. Haidt, «Why universities must choose one telos: truth or social justice», *Heterodox Academy*, 16 de octubre de 2016, <https://heterodoxacademy.org/blog/one-telos-truth-or-social-justice-2/>.

81. American Academy of Arts and Sciences, *Perceptions of Science...*, *op. cit*.

82. Nyhan, «Building a better correction: Three lessons from new research on how to counter misinformation», *Columbia Journalism Review*, 2013, <http://archives.cjr.org/united_states_project/building_a_better_correction_nyhan_new_misperception_research.php>; Nyhan y Reifler, *Misinformation and Fact-Checking: Research Findings from Social Science*, Washington, D. C., New America Foundation, 2012.

83. Willingham, «Critical thinking: Why is it so hard to teach?», *American Educator*, 31, 2007, págs. 8-19, <https://doi.org/10.3200/AEPR.109.4.21-32>.

84. Bond, «Risk school», *Nature*, 461, págs. 1189-1192, 28 de octubre, 2009; Hoffrage, Lindsey, Hertwig y Gigerenzer, «Communicating statistical...», art. cit.; Lilienfeld, Ammirati y Landfield, «Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare?», *Perspectives on Psychological Science*, 4, 2009, págs. 390-398, <https://doi.org/10.1111/j.1745-6924.2009.01144.x>; Mellers, Ungar, Baron *et al*., «Psychological strategies for winning a geopolitical forecasting tournament», *Psychological Science*, 25, 2014, págs. 1106-1115, <https://doi.org/10.1177/0956797614524255>; Morewedge, Yoon, Scopelliti *et al*., «Debiasing decisions: Improved decision making with a single training intervention», *Policy Insights from the Behavioral and Brain Sciences*, 2, 2015, págs. 129-140, <https://doi.org/10.1177/2372732215600886>; Willingham, «Critical thinking...», art. cit.

85. Kahan, Wittlin, Peters *et al*., «The tragedy...», art. cit.; Stanovich, *The Bias That Divides*..., *op. cit*.

86. Ellickson, *Order without Law..., op. cit.*; Ridley, *The Origins of Virtue..., op. cit.*

87. Rauch, *The Constitution...*, *op. cit*.; Sloman y Fernbach, *The Knowledge Illusion...*, *op. cit*.

88. Eisenstein, *The Printing Revolution in Early Modern Europe*, Nueva York, Cambridge University Press, 2012 (2.ª ed.) (trad. cast.: *La revolución de la imprenta en la Edad Moderna europea*, Madrid, Akal, 1994).

89. Kräenbring, Monzon Penza, Gutmann *et al.*, «Accuracy and completeness of drug information in Wikipedia: A comparison with standard textbooks of pharmacology», *PloS ONE*, 9, 2014, e106930, <https://doi.org/10.1371/journal.pone.0106930>.

90. Véase Wikipedia: «List of policies and guidelines», <https://en.wikipedia.org/wiki/Wikipedia:List_of_policies_and_guidelines> y «Five pillars», <https://en.wikipedia.org/wiki/Wikipedia:Five_pillars>.

91. Reforma de los medios sociales: Fox, «Social media: How might it be regulated?», *BBC News*, 12 de noviembre de 2020, <https://www.bbc.com/news/technology-54901083>; Lyttleton, «Social media is determined to slow the spread of conspiracy theories like QAnon. Can they?», *Millennial Source*, 28 de octubre de 2020, <https://themilsource.com/2020/10/28/social-media-determined-to-slow-spread-conspiracy-theories-like-qanoncan-they/>. Algunos análisis tempranos: Pennycook, Cannon y Rand, «Prior exposure increases perceived accuracy of fake news», *Journal of Experimental Psychology: General*, 147, 2018, págs. 1865-1880, <https://doi.org/10.1037/xge0000465>; Pennycook y Rand, «The cognitive science...», art. cit.

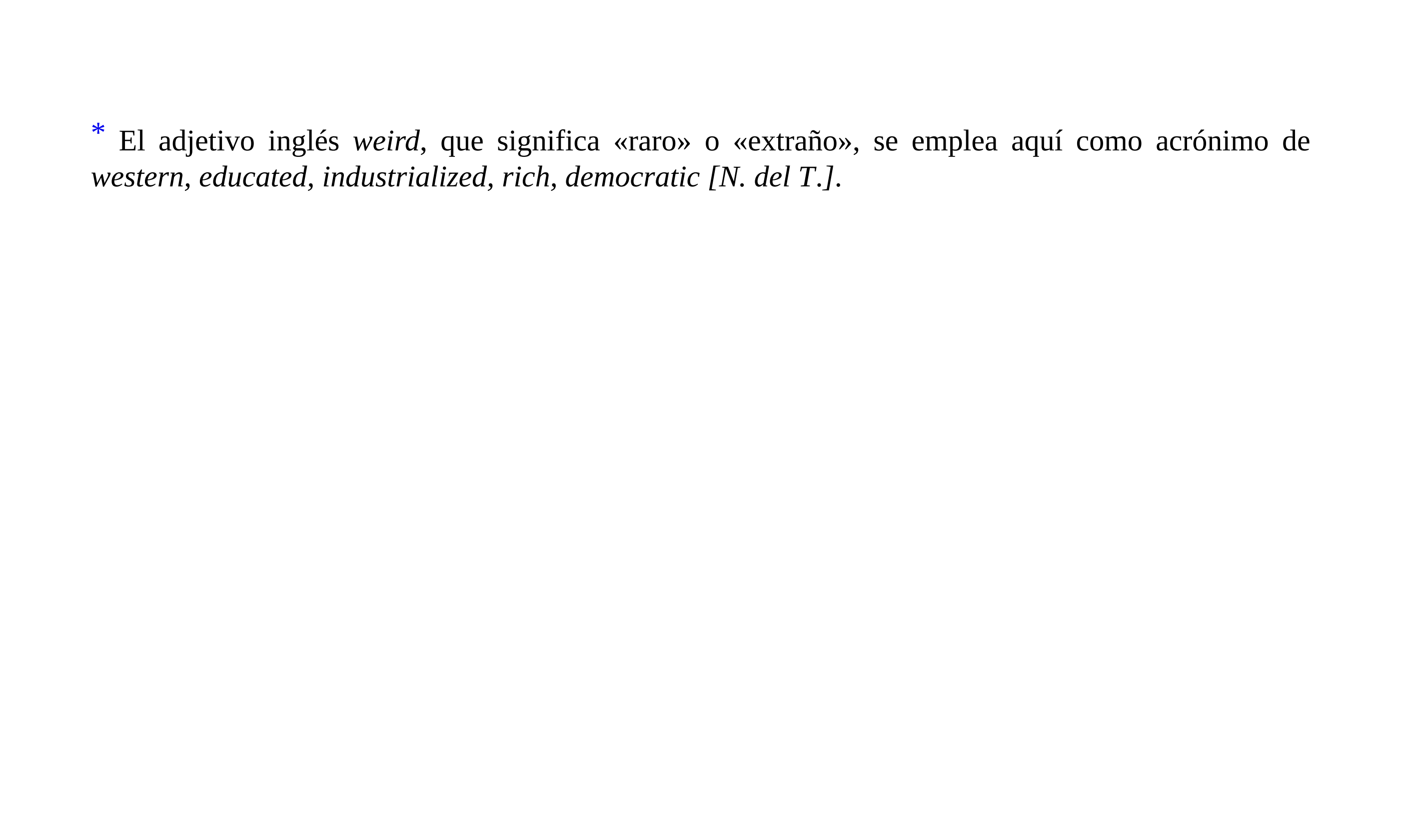
92. Joyner, «Ranking the pundits: A study shows that most national columnists and talking heads are about as accurate as a coin flip», *Outside the Beltway*, 3 de mayo de 2011, <https://www.outsidethebeltway.com/ranking-the-pundits/>; Tetlock, «All it takes to improve forecasting is keep score», Seminarios sobre Pensamiento a Largo Plazo, San Francisco, 23 de noviembre de 2015.

93. Pinker, *Enlightenment Now...*, *op. cit.*, págs. 380-381.

94. Elster, *Deliberative Democracy*, Nueva York, Cambridge University Press, 1998 (trad. cast.: *La democracia deliberativa*, Barcelona, Gedisa, 2001); Fishkin, *When the People Speak: Deliberative Democracy and Public Consultation*, Nueva York, Oxford University Press, 2011.

95. Mercier y Sperber, «Why do humans...», art. cit.

* *«Lay not the flattering unction to your soul» [N. del T.].*



* El adjetivo inglés *weird*, que significa «raro» o «extraño», se emplea aquí como acrónimo de *western, educated, industrialized, rich, democratic [N. del T.]*.

1. Singer, *The Expanding Circle...*, *op. cit.*, pág. 88.

2. Para un análisis incisivo de conflicto versus error como impulsores del progreso humano, véase Alexander, Conflict vs. mistake», *Slate Star Codex*, 2018, <https://slatestarcodex.com/2018/01/24/conflictvsmistake/>.

3. Estos ejemplos se comentan en los capítulos 4-9; véase también Stanovich, «How to think rationally about world problems», *Journal of Intelligence*, 6(2), 2018, <https://doi.org/10.3390/jintelligence6020025>; Stanovich, West y Toplak, *The Rationality Quotient...*, *op. cit*.

4. Stanovich, «How to think...», art. cit.

5. Véase <http://whatstheharm.net/index.html>. Muchos de sus ejemplos están respaldados por informes científicos, enumerados en <http://whatstheharm.net/scientificstudies.html>. Farley dejó de mantener el sitio en torno a 2009, pero refiere esporádicamente ejemplos en su canal de Twitter @WhatsTheHarm, <https://twitter.com/whatstheharm>.

6. Bruine de Bruin, Parker y Fischhoff, «Individual differences in adult decision-making competence», *Journal of Personality and Social Psychology*, 92, 2007, págs. 938-956, <https://doi.org/10.1037/0022-3514.92.5.938>.

7. Ritchie, *Intelligence...*, *op. cit*.

8. Bruine de Bruin, Parker y Fischhoff, «Individual differences...», art. cit. Véase también Parker, Bruine de Bruin, Fischhoff y Weller, «Robustness of decision-making competence: Evidence from two measures and an 11-year longitudinal study», *Journal of Behavioral Decision Making*, 31, 2018, págs. 380-391, <https://doi.org/10.1002/bdm.2059>, para el seguimiento de niños de once años, y Toplak, West y Stanovich, «Real-world correlates of performance on heuristics and biases tasks in a community sample», *Journal of Behavioral Decision Making*, 30, 2017, págs. 541-554, <https://doi.org/10.1002/bdm.1973>, para resultados similares. En 2020, la economista Mattie Toma y yo replicamos el resultado en una encuesta a 157 estudiantes de Harvard matriculados en mi curso Racionalidad (Toma, «Gen Ed 1066 decision-making competence survey», Universidad de Harvard, 2020).

9. Pinker, *The Better Angels...*, *op. cit.*; Pinker, *Enlightenment Now...*, *op. cit*. Conclusiones relacionadas: Kenny, *Getting better: Why Global Development Is Succeeding – And How We Can Improve the World even More*, Nueva York, Basic Books, 2011; Norberg, *Progress: Ten Reasons to Look Forward to the Future*, Londres, Oneworld, 2016 (trad. cast.: *Progreso: diez razones para mirar al futuro con optimismo*, Barcelona, Deusto, 2017); Ridley, *The Rational Optimist: How Prosperity Evolves*, Nueva York, HarperCollins, 2010 (trad. cast.: *El optimista racional: ¿tiene límites la capacidad de progreso de la especie humana?*, Madrid, Taurus, 2011), y los sitios web Our World in Data (<https://ourworldindata.org>) y Human Progress (<https://www.humanprogress.org/>).

10. Roser, Ortiz-Ospina y Ritchie, «Life Expectancy», *Our World in Data*, 2013, <https://ourworldindata.org/life-expectancy>, consultado el 8 de diciembre de 2020; Pinker, *Enlightenment Now...*, *op. cit*., caps. 5-6.

11. Pinker, *Enlightenment Now..., op. cit.*, cap. 7.

12. Roser, «Economic growth», *Our World in Data*, 2016, <https://ourworldindata.org/economic-growth>, consultado el 8 de diciembre de 2020; Pinker, *Enlightenment Now...*, *op. cit*., cap. 8.

13. Pinker, *The Better Angels..., op. cit.*, caps. 5 y 6; Pinker, *Enlightenment Now..., op. cit.*, cap. 11. Conclusiones relacionadas, en Goldstein, *Winning the War on War: The Decline of Armed Conflict Worldwide*, Nueva York, Penguin, 2011; Mueller, *The Stupidity of War: American Foreign Policy and the Case for Complacency*, Nueva York, Cambridge University Press, 2021; Payne, *A History of Force: Exploring the Worldwide Movement against Habits of Coercion, Bloodshed, and Mayhem*, Sandpoint (ID), Lytton, 2004.

14. Una hoja de ruta para solucionar la crisis climática, en GoldsteinRose, *The 100 % Solution: A Plan for Solving Climate Change*, Nueva York, Melville House, 2020.

15. Pinker, *The Better Angels...*, *op. cit.*, caps. 4, 7; Pinker, *Enlightenment Now...*, *op. cit.*, cap. 15. Conclusiones relacionadas en Appiah, *The Honor Code: How Moral Revolutions Happen*, Nueva York, W. W. Norton, 2010; Grayling, *Toward the Light...*, *op. cit.*; Hunt, *Inventing Human Rights: A History*, Nueva York, W. W. Norton, 2007 (trad. cast.: *La invención de los derechos humanos*, Barcelona, Tusquets, 2009); Payne, *A History of Force...*, *op. cit.*; Shermer, *The Moral Arc: How Science and Reason Lead Humanity Toward Truth*, *Justice*, *And Freedom*, Nueva York, Henry Holt, 2015; Singer, *The Expanding Circle...*, *op. cit*.

16. Alexander, «Conflict vs. mistake», art. cit.

17. Pinker, *The Better Angels...*, *op. cit*., cap. 4; véase también Appiah, *The Honor Code...*, *op. cit*.; Grayling, *Toward the Light...*, *op. cit*.; Hunt, *Inventing Human Rights...*, *op. cit*.; Payne, *A History of Force...*, *op. cit*.

18. Welzel, *Freedom Rising...*, *op. cit.*, pág. 122; véase Pinker, *Enlightenment Now...*, *op. cit.*, pág. 228 y nota 45, y págs. 233-235 y nota 8.

19. *Concerning Heretics, Whether They are to be Persecuted*, citado en Grayling, *Toward the Light...*, *op. cit*., págs. 53-54.

20. Mueller, *The Stupidity of War...*, *op. cit*.

21. Erasmo, D., *The Complaint of Peace: To Which Is Added, Antipolemus; Or, the Plea of Reason, Religion, and Humanity, against War*, Miami, Florida, HardPress, 2017 [1517] (trad. cast.:, *Lamento de la paz*, Barcelona, Acantilado, 2020).

22. Beccaria, *On Crimes and Punishments and Other Writings*, Nueva York, Cambridge University Press, 2010 [1764] (trad. cast.: *De los delitos y las penas*, Madrid, Biblioteca Nueva, 2018); mi mezcla procede de dos traducciones.

23. Pinker, *Enlightenment Now...*, *op. cit.*, págs. 211-213.

24. Bentham y Crompton, «Offences against one’s self: Paederasty (part I)», *Journal of Homosexuality*, 3, 1978 [1785], págs. 389-405, <https://doi.org/10.1300/J082v03n04_07>.

25. Bentham, *An Introduction to the Principles of Morals and Legislation*, 1789, <https://www.econlib.org/library/Bentham/bnthPML.html> (trad. cast.: *Introducción a los principios de la moral y la legislación*, Barcelona, Marbot, 2015), cap. 19.

26. Singer, *The Expanding Circle...*, *op. cit*.

27. Davis, *Slavery and Human Progress*, Nueva York, Oxford University Press, 1984.

28. Locke, *The Second Treatise of Civil Government*, Peterborough, Broadview Press, 2015 [1689] (trad. cast.: *Segundo tratado sobre el Gobierno civil*, Madrid, Alianza, 2014)., tratado II, cap. VI, sec. 61.

29. *Ibidem*, cap. IV, sec. 22.

30. Astell, *Some Reflections upon Marriage: To Which Is Added a Preface, in Answer to Some Objections*, Farmington Hills (MI), Gale ECCO, 2010 [1730].

31. Wollstonecraft, *A Vindication of the Rights of Woman with Structures on Political and Moral Subjects*, Nueva York, Cambridge University Press, 1995 [1792] (trad. cast.: *Vindicación de los derechos de la mujer*, Barcelona, Penguin Clásicos, 2020).

32. Douglass, «What to the slave is the Fourth of July?», en P. S. Foner (comp.), *Frederick Douglass: Selected speeches and writings*, Chicago, Lawrence Hill, 1999 [1852].

Racionalidad
Steven Pinker

Título original: *Rationality: What It Is, Why It Seems Scarce, Why It Matters*
Publicado originalmente en inglés por Viking, un sello editorial de Penguin Random House LLC.

Paidós es un sello editorial de Editorial Planeta, S. A.
Av. Diagonal, 662-664, 08034 Barcelona (España)
www.planetadelibros.com

Primera edición en libro electrónico (epub): octubre de 2021

ISBN: 978-84-493-3880-9 (epub)

Conversión a libro electrónico: Newcomlab, S. L. L.
www.newcomlab.com

¡Encuentra aquí tu próxima lectura!



¡Síguenos en redes sociales!